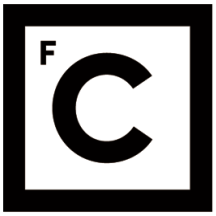
UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS



Ciências
ULisboa

Comparative genomics in *Brucella suis*: from intra-specific and inter-specific distinctive features to diagnostic molecular markers

Doutoramento em Biologia
Especialidade de Microbiologia

Ana Cristina Ribeiro Alves Ferreira Inácio

Tese orientada por:

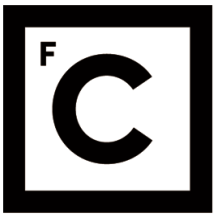
Professor Doutor Rogério Paulo de Andrade Tenreiro
Doutora Maria Inácia Aleixo Vacas de Carvalho Corrêa de Sá

Documento especialmente elaborado para a obtenção do grau de doutor

2017

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS



Ciências
ULisboa

Comparative genomics in *Brucella suis*: from intra-specific and inter-specific distinctive features to diagnostic molecular markers

Doutoramento em Biologia
Especialidade de Microbiologia

Ana Cristina Ribeiro Alves Ferreira Inácio

Tese orientada por:

Professor Doutor Rogério Paulo de Andrade Tenreiro
Doutora Maria Inácia Aleixo Vacas de Carvalho Corrêa de Sá

Júri:

Presidente:

- Doutor José Manuel Gonçalves Barroso

Vogais:

- Doutor João Paulo dos Santos Gomes
- Doutor Albano Gonçalo Beja-Pereira
- Doutora Maria Inácia Vacas de Carvalho Corrêa de Sá
- Doutora Lélia Mariana Marcão Chambel
- Doutor Ricardo Pedro Moreira Dias

Documento especialmente elaborado para a obtenção do grau de doutor

2017

It is not the strongest of the species that survive,

nor the most intelligent, but the one most adaptable to change.

Wrongly attributed to Charles Darwin.
Leon Megginson, 1964.

Acknowledgments

First of all I would like to thank my supervisors, Professor Rogério Tenreiro and Doctor Maria Inácia Corrêa de Sá, for giving me the opportunity to work on this thesis, as well as for their friendly and good-humored mentoring, constant availability and guidance. I am very grateful for their support throughout my studies and doubts. Professor Rogério, thank you for receiving me so well at BioISI Research group M&B, and for always trying to meet the best conditions so that I could achieve my goals. Doctor Maria Inácia, I would like to express my admiration and gratitude for you and for all the opportunities of valorization and scientific training you have given me during these last 20 years. Thank you for your interest and most of all, thank you for all the advices (often personal) and words of encouragement.

A special acknowledge to Doctor Ricardo Dias from the BioISI Research group M&B for open the doors for me on the whole-genome sequencing field, and for all the support and guidance in the comparative genomic studies. Thank you for sharing knowledge and suggestions, constant availability, constructive criticism and for your friendship, having always respected the way I work. Without your support this journey would have been much more difficult!

Thanks to all the members of the M&B group (the "bugworkers family") who in one way or another contributed for the success of this PhD work. Thanks for making me feel as a bugworker member, for all the good moments and fruitful scientific (or not) discussions. A special thanks to Doctor Lélia Chambel for being always available to assist me with BioNumerics and for all the precious advices and friendship. I also want to thank the colleagues that help me out with the Sanger sequencing in the beginning of this work.

I would like to thank my superiors, Doctor Miguel Fevereiro and Doctor Ana Botelho, and all my colleagues from the Instituto Nacional de Investigação Agrária e Veterinária (INIAV, IP) that, directly or indirectly, contributed to the accomplishment of this work. In particular, special thanks to my colleagues from the brucellosis sector, Isabel Travassos Dias and Regina Cardoso, and to my dear friends Helga Waap, Cristina Ferreira and Alice Geraldes, for their friendship, support in the most difficult moments and for always believe that I could do this.

I'm very grateful to my friends Lúcia, Teresa and Marta for making part of my life, and for always believing in my capability to go further.

I wish to thank my parents, in-laws, brothers, sisters-in-law and nephews, for being always around and for encouraging me to go on. A special thanks to my parents and in-laws for all the support. Without you it would be much more difficult.

Finally, I want to deeply thank my dearest husband, Daniel, for always being by my side, for the patience and above of all, for all the love. You will be in my heart forever! To my favorite persons

in the world, my children, Rodrigo and Maria, I just want to say that I love you. This thesis is dedicated to you!

Support from Biosystems and Integrative Sciences Institute (BioISI, FCT/UID/Multi/04046/2013), FCT/MCTES/PIDDAC (Portugal) and FCT project PTDC/CVT/104050/2008 is also acknowledged.

Table of Contents

Abstract

Brucella suis is divided into five biovars of which only biovars 1, 2 and 3 infect Suidae. Biovars 1 and 3 cause severe disease in humans and are mainly prevalent in South America and South-East Asia. In contrast, biovar 2 has been rarely isolated from humans, and its zoonotic role is questioned. Currently, *B. suis* biovar 2 is the unique biovar isolated in Portugal and Spain, representing an emerging disease in domestic swine throughout Europe, being associated with the increase of extensive pig farms and the high density of infected wild boars (*Sus scrofa*).

In the present work the genetic structure of a collection of *B. suis* strains was characterized. Molecular fingerprinting with restriction enzymes and Variable Number Tandem Repeats (VNTR) showed differences between strains from Iberian Peninsula and others from Central Europe. The majority of strains isolated in Portugal and Spain share specific molecular characteristics establishing an Iberian clonal lineage. However, strains isolated from wild boars in the North-East region of Spain were similar to those isolated in other Central-European countries.

In order to deeply understand genomic differences between Iberian and Central-European clones and further unveil *B. suis* pan-genome and evolutive history, the genomes of five *B. suis* biovar 2 strains isolated from wild boars and representative of both clonal lineages circulating in Iberian Peninsula were sequenced and compared with 18 publicly available sequenced genomes from eight of the 12 recognized *Brucella* species. This full genome comparative analysis showed that *Brucella* is a highly conserved genus with two chromosomes and apparently slow evolutionary rates at species level. Nevertheless, *B. suis* biovar 2 strains from Iberian clonal lineage can be differentiated from those from Central-European clonal lineage not only by the presence of one large inversion in Chromosome I but also by a number of specific SNPs, deletions and insertions. Furthermore, 10 different target-PCRs protocols were established and validated for the differentiation of strains from a specific clonal lineage, being useful to be used as epidemiological markers for future investigations. The mutational enrichment of Iberian lineage was associated to genes encoding membrane proteins described with potential of interaction with external stimulus, as well as to genes with impact on the metabolism of the pathogen. The genomic specialization and local adaptation of *B. suis* biovar 2 strains establish an Iberian ecovar, raising an important question regarding the mechanisms responsible for putative tropisms as response to adaptation to a specific host and/or pathobiological conditions. Future work should be done to better understand the consequences of these disarrangements and their impact in pathogenicity or virulence in wide range of hosts, including man.

Keywords: *B. suis* biovar 2; Whole-genome sequencing; Optical mapping; Comparative genomics; Genomic specialization.

Resumo

A brucelose em suínos é uma infecção causada por *Brucella suis*. Esta espécie apresenta cinco biovares, sendo a infecção em suínos causada apenas pelos biovares 1, 2 e 3. Os biovares 1 e 3 são patogénicos para o homem e prevalecem sobretudo na América do Sul e Sudeste da Ásia. O biovar 2 não é patogénico para o homem e é enzoótico nas populações de javalis (*Sus scrofa*) e lebres (*Lepus europaeus*) do Norte, Centro e Sudeste da Europa, sendo estes animais silváticos os agentes transmissores da infeção aos suínos. De facto, a brucelose devida ao biovar 2 representa uma doença emergente em suínos domésticos em toda a Europa e está associada ao aumento de explorações extensivas e à alta densidade de javalis infetados, representando um perigo importante especialmente para a população de porcos ibéricos criados em sistemas extensivos.

Na Península Ibérica, o biovar 2 é o único biovar isolado em porcos e javalis. Apesar da escassez de estudos para compreender as relações epidemiológicas e a filogenia deste agente patogénico, a tipificação molecular baseada em polimorfismos no tamanho de fragmentos de hidrólise com enzimas de restrição (RFLPs) de genes codificantes de proteínas da membrana externa (*Omp2a*, *Omp2b*, *Omp31*) e em variações no número de repetições em tandem em regiões micro- ou minissatélites do genoma (VNTRs) mostraram diferenças entre as estirpes da Península Ibérica e as da Europa Central. Assim, o trabalho desenvolvido ao longo desta tese visou aumentar o conhecimento das estirpes de *B. suis* biovar 2 que circulam na Europa e, em particular, em Portugal, bem como evidenciar variações genómicas associadas a estirpes de origem geográfica específica.

Em Portugal, o diagnóstico bacteriológico da brucelose é apenas realizado no Laboratório de Bacteriologia e Micologia da Unidade Estratégica de Produção e Saúde Animal do Instituto Nacional de Investigação Agrária e Veterinária (INIAV, IP). No entanto, este diagnóstico está especialmente focado no isolamento de *B. abortus* e *B. melitensis* na sequência das campanhas de controlo e erradicação da brucelose em bovinos e pequenos ruminantes. Relativamente aos suínos, não é efetuada qualquer vigilância sistemática, desconhecendo-se a prevalência da infeção em Portugal.

Os estudos efetuados neste trabalho visaram alargar o conhecimento sobre a diversidade genética das estirpes de *B. suis* biovar 2 que circulam em Portugal. Inicialmente foi desenvolvido e avaliado um meio seletivo, LNIV-M, que foi formulado utilizando antibióticos menos inibitórios, permitindo aumentar a sensibilidade do diagnóstico bacteriológico e aumentar o número de isolados de *B. suis*. Uma vez que o conhecimento das estirpes predominantes é um pré-requisito para qualquer estudo epidemiológico, foram aplicados diferentes métodos moleculares a uma coleção de isolados de *B. suis*, que incluiu não só os isolados de Portugal como também de outros países europeus. Foram identificados cinco perfis de restrição (haplótipos) entre os isolados de biovar 2, com dois haplótipos específicos restritos a Portugal e Espanha (2d e 2e). Os três haplótipos restantes (2a, 2b e 2c),

encontram-se disseminados pela Europa (exceto na Península Ibérica). A diversidade genética da população de *B. suis* foi determinada pela análise do número variável de repetições em tandem em múltiplos locais do genoma, vulgarmente designada por MLVA. O ensaio de MLVA aplicado à genotipificação de estirpes de *Brucella* spp. utiliza um conjunto de 16 VNTRs que estão distribuídos por três painéis: painel 1, composto por oito minissatélites, painel 2A e painel 2B, compostos respetivamente por três e cinco microssatélites. A análise alocou os isolados de biovar 2 em dois *clusters* de acordo com as suas origens geográficas e haplótipos, definindo assim uma linhagem clonal ibérica (Portugal e Espanha) e uma linhagem clonal Central-Europeia. A análise de 350 estirpes adicionais de todas as biovariedades de *B. suis* revelou ainda uma elevada divergência genética entre as estirpes com base nos seus hospedeiros, realçando a estreita relação entre as estirpes de suínos, javalis e lebres. Além de corroborar a existência das duas linhagens clonais de biovar 2, os resultados obtidos sugerem que a evolução da linhagem clonal Ibérica ocorre devido a um evento de especiação alopátrica.

Para comparar a estrutura genómica e avaliar a diversidade genética entre as estirpes de *B. suis* biovar 2, foram construídos os mapas ópticos de cinco estirpes de campo isoladas de javalis e representativas das linhagens clonal Ibérica (PT09143, PT09172, Bs143CITA) e Central-Europeia (Bs364CITA, Bs396CITA), bem como da estirpe de referência *B. suis* biovar 2 ATCC 23445 (linhagem Central-Europeia, origem: Dinamarca), utilizando a tecnologia *Optical Mapping* (mapeamento óptico). Cada estirpe apresentou um perfil único de restrição, de 228 a 232 fragmentos, distribuídos pelos dois cromossomas, tendo-se observado baixa divergência no cromossoma II (1,6%) relativamente ao cromossoma I (2,4%). O mapeamento óptico revelou no cromossoma I a presença de um evento de inserção-deleção (INDEL, 3,5 kb) específico da estirpe de referência ATCC 23445, e uma grande inversão (944 kb) exclusiva da linhagem clonal Ibérica.

Além de validar a existência de uma inversão na linhagem clonal Ibérica, reforçando a plasticidade genómica de *Brucella* spp., os mapas ópticos permitiram também determinar o posicionamento e a orientação das sequências consenso (*contigs*), agilizando o processo de montagem e finalização dos genomas. Assim, as sequências genómicas completas e anotadas, das cinco estirpes de *B. suis* biovar 2 referidas anteriormente, foram prontamente obtidas utilizando-se uma estratégia que combinou: (1) a sequenciação das estirpes, utilizando a plataforma de sequenciação de nova geração (NGS), Illumina HiSeq 2000; (2) a re-sequenciação, por Sanger, das regiões de baixa qualidade ou contendo bases ambíguas; (3) a montagem *de novo* e finalização dos genomas guiada por *Optical mapping*. À semelhança do genoma da estirpe de referência de *B. suis* biovar 2, os cinco genomas sequenciados neste trabalho são compostos por dois cromossomas circulares (Chr I e Chr II) com aproximadamente 1,93 e 1,40 Mb. Em média o teor GC (%) dos Chr I e Chr II foi de 57,1% e 57,3%, respetivamente. Os cinco genomas contêm três operões idênticos de genes de RNA ribossomal (rRNA), um localizado no Chr I e dois no Chr II. Todos os genomas possuem 54 genes de RNA de transferência

(tRNA). O número de regiões codificantes (CDS) foi estimado tendo por base a homologia e a identificação dos domínios funcionais das proteínas, que variaram de 3 014 (PT09143) a 3 027 (Bs396CITA) e o número de pseudogenes entre 87 (Bs396CITA) e 91 (Bs364CITA).

A fim de melhor compreender os mecanismos envolvidos na evolução e especialização das linhagens Ibéricas, foi realizada uma análise genómica comparativa entre os cinco genomas de *B. suis* biovar 2 e os 18 genomas disponíveis nas bases de dado públicas, incluindo oito espécies de *Brucella*. A análise efetuada confirmou que o género *Brucella* é altamente conservado, com dois cromossomas e taxas de evolução aparentemente lentas a nível de espécie. No entanto, as estirpes de *B. suis* biovar 2 da linhagem clonal Ibérica foram diferenciadas da Central-Europeia não só pela presença de uma grande inversão no Chr I, mas também por um conjunto de polimorfismos de base única (SNPs) e deleções-inserções (INDELs) específicas. Estes resultados permitiram desenvolver e validar diferentes protocolos de PCR para 10 regiões polimórficas com potencial epidemiológico para a diferenciação das estirpes de uma linhagem clonal específica. Foi ainda observado que o enriquecimento mutacional da linhagem Ibérica está associado a genes que codificam proteínas de membrana, bem como a genes com impacto no metabolismo deste agente patogénico. No entanto, outros estudos são necessários para compreender melhor as consequências destas variações genómicas e o seu impacto na patogenicidade ou virulência numa ampla gama de hospedeiros, incluindo o Homem.

Em conclusão, com este trabalho foi demonstrado que existem duas linhagens clonais de *B. suis* biovar 2 a circular na Europa, sendo que a especialização genómica e adaptação local das estirpes pertencentes à linhagem clonal Ibérica estabelecem um ecovar Ibérico, levantando uma importante questão sobre os mecanismos responsáveis por tropismos putativos como resposta à adaptação a um hospedeiro específico e/ou condições patobiológicas.

Palavras-chave: *B. suis* biovar 2; *Optical Mapping*; Sequenciação de Nova geração; Genómica Comparativa; Filogenómica.

Table Index

Figure Index

Supplementary Material Index

List of abbreviations

ATCC	American Type Culture Collection
BAB	Blood agar base
BioISI	Biosystems & Integrative Sciences Institute
Blast	Basic Local Alignment Search Tool
BMB	*Brucella* medium base
CDS	Coding DNA sequence
CEG	Concentration enabling growth
CFU	Colony-forming unit
Chr	Chromosome
DNA	Deoxyribonucleic acid
Dntp	Deoxyribonucleotide triphosphate
FAR	Farrell's medium
GC	GC medium
GO	Gene Ontology
GT	Genotype
HGDI	Hunter-Gaston diversity index
INDEL	Insertion-deletion event
INIAV	Instituto Nacional de Investigação Agrária e Veterinária
IS	Insertion sequence
KEGG	Kyoto Encyclopedia of Genes and Genomes
LPS	Lipopolysaccharide
MIC	Minimal inhibitory concentration
ML	Maximum likelihood
MLSA	Multilocus sequencing analysis
MLST	Multilocus sequencing typing
MLVA	Multilocus variable number tandem repeat analysis
MST	Minimum spanning tree
MTM	Modified Thayer-Martin's medium
NCBI	National Center for Biotechnology Information
NJ	Neighbour-Joining
NCTC	National Collection of Type Cultures
NGS	Next generation sequencing
OM	Optical map
OMP	Outer membrane protein
ORF	Open reading frame
PATRIC	Pathosystems Resource Integration Center
PCR	Polymerase chain reaction
PHAST	Phage search tool
PM	Plommet medium
RAST	Rapid annotations using subsystems technology

RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
tRNA	Transfer RNA
R-LPS	Rough-LPS
S-LPS	Smooth- LPS
SNP	Single nucleotide polymorphism
TSA	Trypticase soy agar
UPGMA	Unweighted pair group method with arithmetic average
VNTR	Variable number tandem repeat
WG	Whole-genome
WGM	Whole-genome mapping
WG-INDEL	Whole-genome insertion-deletion events distribution
WG-MSA	Whole-genome multiple sequence alignment
WG-SNP	Whole-genome single nucleotide polymorphisms distribution

Research papers integrated into this thesis:

- Ferreira, A.C., Corrêa de Sá, M.I., Dias, R. and Tenreiro, R. (2016). Whole-genome mapping reveals a large genetic inversion on Iberian *Brucella suis* biovar 2 strains. *Veterinary Microbiology*, 192: 220–225. doi:10.1016/j.vetmic.2016.07.024.
- Ferreira, A.C., Corrêa de Sá, M.I., Tenreiro, R. and Dias, R. (2014). Complete genome sequences of three Iberian *Brucella* suis biovar 2 strains isolated from wild boars. *Genome Announcements*. 2(4):e00618-14. doi:10.1128/genomeA.00618-14.
- Ferreira, A.C., Corrêa de Sá, M.I., Tenreiro, R. and Dias, R. (2014). Complete genome sequences of two Central-European *Brucella* suis bv. 2 haplotype 2c strains isolated from wild boars. *Genome Announcements*. 2(4):e00686-14. doi:10.1128/genomeA.00686-14.
- Ferreira, A.C., Almendra, C., Cardoso, R., Pereira, M.S., Beja-Pereira, A., Luikart, G., Corrêa de Sá, M.I. (2012). Development and evaluation of a selective medium for *Brucella* suis. *Res. Vet. Sci*. 93(2): 565-567.

Submitted research papers:

- Ferreira, A.C., Tenreiro, R., Dias, R. and Corrêa de Sá, M.I. (2016). Genetic diversity of *Brucella suis* biovar 2 strains circulating in Europe. *Submitted to Veterinary Microbiology*.
- Ferreira, A.C., Tenreiro, R., Corrêa de Sá, M.I. and Dias, R. (2017). Evolution and genomic specialization of *Brucella suis* biovar 2 Iberian Lineages. *Submitted to BMC Genomics.*

Chapter 1

General Introduction

1.1. The Genus *Brucella*: a taxonomic and phylogenetic overview

In the year of 1887 in Malta, with the isolation and identification of a bacterium by David Bruce and Themistocles Zammit, now known as *Brucella melitensis* (**Figure 1.1**), began the history of Brucellosis, one of the most extended bacterial zoonosis at a global level and a complex infection of animals and humans with a worldwide impact. This bacteria was first isolated in Malta from the spleens of soldiers with fatal cases of brucellosis, also known as undulant fever or Malta fever. The genus *Brucella*, created in 1920 by Meyer and Shaw (De Ley *et al.*, 1987), belongs to the family Brucellaceae within the order Rhizobiales of the class Alphaproteobacteria, which is one of the largest and most diverse groups within the phylum Proteobacteria (Scholz *et al.*, 2012; Godfroid *et al.*, 2011; Ficht, 2010; Audic *et al.*, 2009; Bergey and Holt, 1994). The order Rhizobiales includes a variety of bacteria strategically important for their diversity in function and in niche occupancy, including animal intra or pericellular pathogens (*Afipia*, *Anaplasma, Bartonella, Brucella*, *Erlichia, Ricketsia*), opportunistic animal pathogens (such as *Ochrobactrum*), plant pathogens (*e.g. Agrobacterium*) and several plant endosymbionts (Carvalho *et al.*, 2010; Velasco *et al.*, 1998). Phylogenetic reconstructions based on whole-genome sequences have confirmed the evolutionary proximity between *Brucella* genus and members of the subgroup of the alpha-2 Proteobacteria, that includes soil organisms (*e.g. Ochrobactrum* spp.), plant symbionts (*e.g. Rhizobium* spp.) and phytopathogens (*e.g. Agrobacterium* spp.), and identified the *Ochrobactrum*, a soil living facultative human pathogen, as the most closely related genus (Bohlin *et al.*, 2010; Whatmore *et al.*, 2009; Wattam *et al.*, 2009; Scholz *et al.*, 2008b).

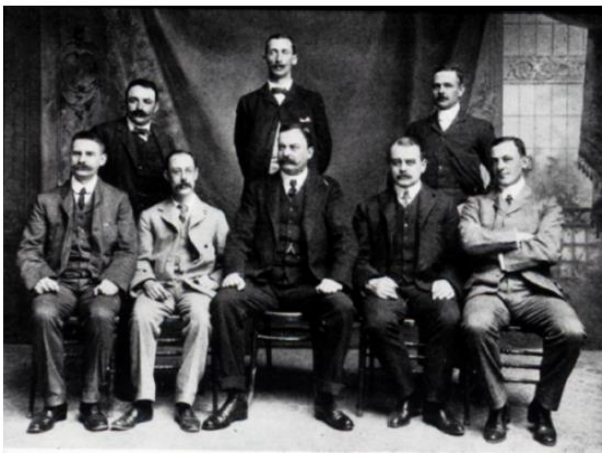


Figure 1.1. Mediterranean fever Commission (MFC) in 1904. *Standing*: Dr. T. Zammit; Capt. Crawford Kennedy; Major J.C. Weir; *Seated*: Major J.G. McNought; Dr. J.W.H. Eyre; Col. David Bruce.; Major T. McCulloch; Staff Surgeon E.M.A. Clayton.

Source:http://wellcomeimages.org/indexplus/obf_images/a5/20/95da16f04714583a1a503321854a.jpg

The genus *Brucella* was created with two species, *B. abortus* and *B. melitensis*, which preferential hosts are cattle and small ruminants (sheep and goats), respectively. *B. suis* was isolated for the first time from aborted pigs fetus in Europe in 1909, and after in the United States. For many years, it was believed that the agent was a highly pathogenic variant of *B. abortus* but, in 1929, *B. suis* was finally considered a separate species. In the following years, new species were added to the genus: *B. ovis* isolated from sheep (1956), *B. neotomae* (1957) from the desert wood rat, and *B. canis* (1968) from dogs. Like the first two species, they were exclusively classified and characterized on the basis of their phenotype and host preference (Alton, 1990). The three major species in terms of disease and economic impact for man, *B. melitensis*, *B. abortus* and *B. suis* are further divided into biovars based on a range of phenotypic and serological characteristics: *B. melitensis* with 3 biovars, *B. abortus* with 8 biovars, and *B. suis* with 5 biovars. Since the genetic homogeneity of these "classical" species seems to support the hypothesis of a monospecific genus, *B. melitensis* has been proposed as the sole representative species with different biotypes (Verger *et al.*, 1985). This classification was accepted by the Sub-committee on the Taxonomy of *Brucella* in 1986. However, the host range was a long-recognized biological criterion and the presence of species specific markers in outer membrane protein genes and in other genes showed that *B. melitensis*, *B. abortus*, *B. suis*, *B. ovis*, *B. canis* and *B. neotomae* were not mere pathovars (or nomenspecies) but biologically meaningful species. Consequently, in 2003, the Sub-committee approved to return to the multi-nomen species classification with the readoption of the "classical" species with a series of biovars (Osterman and Moriyón, 2006; Moreno *et al.*, 2002).

Since 2007, more species were included in the genus: *B. ceti* and *B. pinnipedialis,* isolated from marine mammal (Foster *et al.*, 2007), *B. microti* from voles (Scholz *et al.*, 2008a), *B. inopinata* from an inflamed breast implant of a 71 year-old patient in USA (Scholz *et al.*, 2010) and, more recently, *B. papionis*, isolated from baboons (Whatmore *et al.*, 2014) and *B. vulpis*, isolated from foxes (Scholz *et al.*, 2016). The natural reservoir of *B. inopinata* remains unclear (Eisenberg *et al.*, 2016). Like most *Brucella* species, *B. ceti*, *B. pinnipedialis* and *B. papionis* are fastidious and slow growing pathogens, with limited metabolic activity; in contrast, *B. microti, B. inopinata* and *B. vulpis*, are fast growing bacteria with a biochemical profile that resemble the members of the genus *Ochrobactrum* (Hammerl *et al.*, 2016; Mühldorfer *et al.*, 2016; Scholz *et al.*, 2016, 2008a; Al Dahouk *et al.*, 2010). Other "atypical" *Brucella* strains have been isolated from diverse animal sources such as wild rodents (Tiller *et al.*, 2010), frogs (Eisenberg *et al.*, 2012) and fish (Eisenberg *et al.*, 2016) that will likely to be proposed as new species in the future. Despite of *Brucella* spp. difference in host affinity, they display very similar pathogenic behavior, while varying in virulence. The preferential hosts and the pathogenicity for humans of the 12 recognized *Brucella* species are described in **Table 1.1**.

Table 1.1. *Brucella* species, preferred host(s) and pathogenicity for humans.

***Brucella* species**	**Biovars**	**Preferential host(s)**	**Pathogenicity for humans**
B. melitensis	1–3	Sheep, goat	High
B. abortus	1–7, 9[1]	Cattle	High
B. suis	1, 3	Swine	High
	2	Wild boar, hare	Low or no pathogenicity
	4	Reindeer, caribou	High
	5	Rodents	No
B. neotomae	-	Rodents	No
B. ovis	-	Ram	No
B. canis	-	Dog	Moderate
B. ceti	-	Cetaceans	Unknown
B. pinnipedialis	-	Pinnipeds	Unknown
B. microti	-	Soil, vole, fox	Unknown
B. inopinata	-	Unknown	High
B. papionis	-	Baboons	Unknown
B. vulpis	-	Foxes	Unknown

[1] *B. abortus* biovar 8 was deleted by the *Brucella* Subcommittee in 1978 (Osterman and Moriyón, 2006)

All *Brucella* species have identical 16S ribosomal (r)RNA- and recA gene sequences, and are almost identical in the majority of housekeeping genes (Whatmore *et al.*, 2009, 2007; Scholz *et al.*, 2008b; Gee *et al.*, 2004). The current understanding of *Brucella* spp. phylogeny has been elucidated by several methods such as Multilocus Variable Number Tandem Repeat Analysis (MLVA, Scholz & Vergnaud, 2013; Al Dahouk *et al.*, 2012; Le Flèche *et al.*, 2006), Multi*locus* Sequencing Typing (MLST) and MLS Analysis (MLSA) (Whatmore *et al.*, 2016, 2009, 2007), microarray studies (Foster *et al.*, 2012; Bohlin *et al.*, 2010; Rajashekara *et al.*, 2004), and Single Nucleotide Polymorphisms (SNP, Wattam *et al.*, 2014, 2012, 2009; Foster *et al.*, 2009, 2008; Chain *et al.*, 2005). The phylogenetic analysis showed that all *Brucella* species are monophyletic and distinct from *Ochrobactrum*. Moreover, most of the analysis referred in the literature distinguish at least six lineages within *Brucella* spp.: *B. abortus-B. melitensis; B. canis-B. suis*, *B. ovis*, *B. ceti-B. pinnipedialis*, *B. neotomae*, and *B. microti* (**Figure 1.2**). From these, *B. abortus* and *B. melitensis* are the most closely related, and a close relationship has also been detected between *B. canis* and *B. suis* biovars 3 and 4, while *B. suis* biovars 1, 2 and 5 were allocated to different branches. *B. neotomae* and *B. ovis* demonstrate greater divergence levels from

other *Brucella* species (Olsen & Palmer, 2014; Wattan *et al.*, 2014; Scholz & Vergnaud, 2013). Few studies have been made that include the novel species *B. inopinata, B. papionis* and *B. vulpis*. Nevertheless, a recent work confirms the abovementioned lineages and place *B. papionis* more closely related to *B. ovis.* The "atypical" *B. inopinata* and *B. vulpis* are placed in two separate branches (Whatmore *et al.*, 2016).

1.2. The *Brucella* species: approaches on diversity and epidemiology

The accessibility of whole-genome sequence data opened the way for comprehensive molecular analyses and subsequent development of molecular typing tools that allow identification and differentiation of *Brucella* at the species, biovar and individual strain level (Scholz & Vergnaud, 2013). Since the 90's, several PCR-based methods have been developed and implemented in diagnostic laboratories to confirm pure cultures of brucellae and differentiate among *Brucella* species and biovars and vaccine strains (Whatmore *et al.*, 2016; Scholz & Vergnaud, 2013; López-Goñi *et al.*, 2011; Yu & Nielsen, 2010; Mayer-Scholl *et al.*, 2010). Moreover, considering that distinguishing individual bacterial lineages within a species is the basis of infectious disease epidemiology, several PCR-based genotyping tools have been developed, including enterobacterial repetitive intergenic consensus sequence PCR (ERIC-PCR), repetitive intergenic palindromic sequence PCR (REP-PCR), randomly amplified polymorphic DNA PCR (RAPD-PCR) or arbitrarily primed PCR (AP-PCR), and restriction fragment length polymorphism PCR (RFLP-PCR) of the *omp2* (including *omp2a* and *omp2b* genes) and *omp31 loci*. From these, only RFLP-PCR assay targeting the *omp2* (Cloeckaert *et al.*, 1995) or *omp31* (Vizcaíno *et al.*, 1995) *loci* are more commonly used in laboratories. In fact, PCR-RFLP analysis of *omp2a*, *omp2b* and *omp31* genes define five different restriction patterns (haplotypes) for each *B. suis* biovar (1 to 5) and four additional haplotypes are further disclosed within biovar 2 isolates. This method have been useful to differentiate among Portuguese and Spanish isolates from those isolated in other European countries (Muñoz *et al.*, 2010; Garcia-Yoldi *et al.*, 2007; Ferrão-Beck *et al.*, 2006).

As already mentioned, current approaches to study *Brucella* spp. diversity and phylogeny include methods such as MLVA, MLST/MLSA, SNPs analysis as well as whole-genome sequencing, which serves as a robust and unbiased method to resolve intraspecies relationships for closely related species such as *Brucella* spp. (Tan *et al.*, 2015). Nevertheless, the highly discriminatory MLVA or MLSA represent a perfect first-line tools for molecular epidemiological studies within outbreak investigations, and MLSA is also appropriate for phylogenetic reconstructions, owing to the highly clonal evolution of the different species (Whatmore *et al.*, 2016; Allen *et al.*, 2015; Scholz & Vergnaud, 2013).

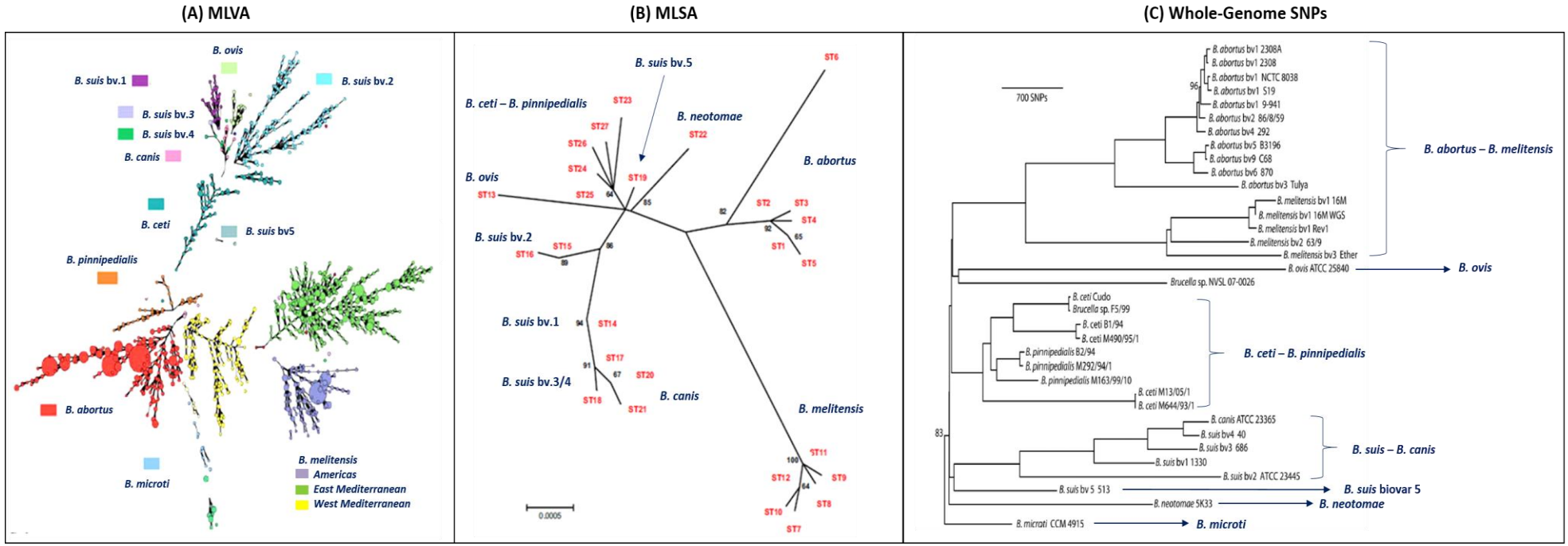


Figure 1.2. Phylogenetic analysis of the *Brucella* species. (A) Minimum spanning tree based on multi*locus* variable number of tandem repeats analysis (MLVA) data from 1,925 isolates. Branch lengths up to three are shown. Adapted from Scholz & Vergnaud, 2013. (B) Unrooted phylogenetic reconstruction of the relationships between sequence types based on multi*locus* sequencing analysis (MLSA). The tree was constructed with the concatenated sequence data of the nine *loci* (4,396 bp) using the neighbour joining approach. Adapted from Whatmore *et al.*, 2007; (C) Phylogenetic tree based on maximum parsimony of the core *Brucella* genomes. The tree was rooted with *B. microti*. All branches have 100% support unless otherwise noted. Adapted from Wattan *et al.*, 2014.

The epidemiological surveillance of animal and human brucellosis has clearly benefited from the appearance and improvement of molecular typing (Valdezate *et al.*, 2010). To date one of the best approaches to obtain valuable discrimination in brucellae is MLVA-16 (using 16 VNTR markers), as it is an easy and cost effective methodology. Since brucellosis has a worldwide distribution, laboratories of different countries try to apply the same genotyping techniques, thus facilitating the exchange of information. In this way, a collaborative public online database based on a MLVA-16 scheme has been built up with the aim of promoting the creation of a global epidemiological map of *Brucella* spp. (http://mlva.u-psud.fr/), where data can be submitted and compared to other published results. MLVA-16 exhibits an intraspecies discriminatory power and is extremely discriminant and highly efficient in differentiating strains within a local outbreak. Most of the epidemiological studies were applied to *B. melitensis* or *B. abortus* due to its important economic and medical concerns. The epidemiological relationship between *B. melitensis* isolates using MLVA-16 assay has shown that three separate clusters are observed when geographical origin is considered: the American, the Eastern Mediterranean, and the Western Mediterranean groups (Valdezate *et al.*, 2010). For instance, in a study using 126 Portuguese isolates, the majority of the isolates (83%) were included in the Eastern Mediterranean group, which also include isolates from Spain and Turkey, and the remainder (17%) was included in the American group (Ferreira *et al.*, 2012). In contrast, most isolates from Italy belong to the Western Mediterranean lineage (Garafolo *et al.*, 2013). Although less variability is found within *B. abortus* populations, MLVA is still a useful tool in ongoing disease surveillance of *B. abortus* outbreaks, especially when combined with accurate epidemiological information on disease tracings, geographical clustering of cases and chronology of infection (Allen *et al.*, 2015). Furthermore, although the validity of *B. abortus* or *B. melitensis* biovars established by classical microbiological methods cannot be confirmed by MLVA clustering, this method can significantly contribute to epidemiological trace-back analysis of human *Brucella* infections and may advance surveillance and control of human brucellosis (Ferreira *et al.*, 2012; Al Dahouk *et al.*, 2007). In addition, it was also showed that MLVA-16 can be an essential assay to guarantee the quality and stability of live anti-bacterial vaccines being produced worldwide, such as the *B. melitensis* REV1 vaccine (Garcia-Yoldi *et al.*, 2007).

Regarding *B. suis* species, recent studies using both MLVA and MLSA have also shown that considerable intraspecies genetic diversity is observed even when considering isolates sharing the same biovar (Whatmore *et al.*, 2016; Kreizinger *et al.*, 2014; Muñoz *et al.*, 2010; Garcia-Yoldi *et al.*, 2007). However, less work have been made in order to better understand the phylogeny and the epidemiologic relationships of this pathogen.

1.3. *Brucella suis* biovar 2 and brucellosis

Within the species *B. suis*, biovars 1, 2 and 3 are the most relevant constituting the main etiological agents involved in swine brucellosis. *B. suis* infection due to biovars 1 and 3 have been reported in several non-natural domestic and wild host animal species, such as cattle, dogs, horses, sheep, reindeer, caribou, hares, red foxes and various murine species. These biovars 1 and 3 are important human pathogens mainly prevalent in South America and South-East Asia (EFSA, 2009). In Europe they have only been reported in Croatia, indicating that may be restricted to this geographic region (Cvetnic *et al.*, 2009; Garcia-Yoldi *et al.*, 2007; Cvetnic *et al.*, 2005).

B. suis biovar 2 has a very specific pathogenicity for suidae and hares but does not infect humans (Olsen & Palmer, 2014). It is usually the causative agent of swine brucellosis in Europe (Godfroid *et al.*, 2011; Muñoz *et al.*, 2010; EFSA, 2009; Garcia-Yoldi *et al.*, 2007) and is widely spread amongst Eurasian wild boar (*Sus scrofa*) and European brown hare (*Lepus europaeus*) populations, which are identified as the potential source of transmission of biovar 2 to outdoor or extensively reared pigs (Olsen & Palmer, 2014; Munoz *et al.*, 2010; Galindo *et al.*, 2010; Cvetnic *et al.*, 2009; Bergagna *et al.*, 2009; Leuenberger *et al.*, 2007). Moreover, it is known that biovar 1, 2 and 3 can be transmitted from swine to cattle, inducing transient seroconversion, which can confound *B. abortus* diagnostic assays (Olsen & Palmer, 2014; Musser *et al.*, 2013).

Brucellae produce abortion and infertility in infected pigs and is easily transmitted in extensive production systems where animals share the same environment and have contacts with wild animals, namely wild boars and hares. The control and eradication of the disease is only possible with well-planned strategies, adapted to the local reality and globally integrated in the production cycle, serving the interest of public and animal health and also of the production economy. Although the infections in wildlife reservoirs are a lesser threat for causing human infection, they can be a source for the reintroduction of infection into domestic livestock, which poses a new challenge to eradication of the disease worldwide (Musser *et al.*, 2013; Plotkin, 2009). The pathogenesis and pathobiology of brucellosis in domestic pigs are likely to be identical to those in wild boar. In a primary infection the bacteria can spread within a few months from one infected animal to more than 50% of animals on the farm. The infection can often reach 70% to 80% of infected animals at the start of the outbreak, and the infected herds manifest a high percentage of abortions, increased neonatal mortality and infertility, with adverse economic consequences (Aparício, 2013). Nevertheless, as there tend to be few -or mild clinical signs, the disease can go unnoticed in infected groups (EFSA, 2009) and abortion may be the only clinical sign observed, usually occurring during the second or third month of gestation (Poester *et al.*, 2013).

Swine brucellosis is considered to be inexistent or of low prevalence in domestic pigs, therefore it is usually excluded from National Surveillance Programs in the European Union. Consequently, diagnosis is only performed on pigs for International Trading purposes, in selected Artificial Insemination Centers or for investigation of suspected cases. However, the increasing number of outbreaks in pig farms all over Europe, and the high prevalence of *B. suis* biovar 2 in wild boars and hares, suggest that swine brucellosis could be an emergent but still unrecognized problem in many countries (Olsen & Palmer, 2014; Godfroid *et al.,* 2013; Munoz *et al.*, 2010; Godfroid & Käsbohrer, 2002).

Occurrence of *B. suis* biovar 2 infection in Portugal

The existence of *B. suis* biovar 2 infection has been confirmed by isolation of the agent from animals belonging to different pig farms in diverse regions of Portugal where sporadic outbreaks occurred. However, no systematic surveillance is carried out and few studies have been performed to evaluate the real status of the infection in Portugal.

There are three recognized autochthonous pig breeds in Portugal reared in extensive production systems, and therefore more susceptible to *B. suis* biovar 2 infection: the Bísaro (produced in Trás-os-Montes), the Alcobaça Spotted ("Malhado de Alcobaça", created in the Centre region), and the Iberian pig ("Porco Alentejano", raised in an unique integrated agricultural-forestry-livestock system in Alentejo, called "montado"). The first documented *B. suis* biovar 2 outbreak in domestic pigs occurred in 1999, where a farm in Alentejo (rearing "Porco Alentejano" breed) was traced back as the source of infection of two other farms (one in the North and other in the Centre, with an intensive production system) due to the introduction of infected males. As the majority of the animals in each farm were serologically positive (near 90%) and there was a high number of abortions, the Veterinary Services decided to slaughter the totality of the animals. After that, between 1999 and 2000, about 12 new outbreaks were identified in Alentejo by the veterinary services. The occurrence of abortions and the increase of serologically positive animals have major economic implications as it prevents the free movement of animals and products. For that reason, the surveillance of brucellosis in domestic pigs as well as in wild boars is an important issue and, when feasible, control and eradication programs should be implemented in the affected areas.

1.4. From the outside-in: the cell envelope and virulence of brucellae

All of the various surface components of a bacterial cell are important in its ecology since they mediate the contact of the bacterium with its environment and consequently supports its own existence and survival in that environment.

Electron microscopy of *Brucella* spp. cells shows the classical structure of the Gram-negative cell envelope with an outer-membrane (OM) of 6.5–8.0 nm and an inner membrane (IM) of similar thickness, both separated by a periplasmic space (PS) (Moreno & Moriyón, 2006). The IM is a phospholipid bilayer that contains proteins involved in subtract transport and other metabolic processes. The PS encloses a peptidoglycan (PG) mesh layer and some periplasmic soluble components, such as cyclic β-glucans, and proteins. The OM is the most external layer of the cell constituting an important barrier for survival in hostile environments and an accessible target for the interaction of bacterial pathogens with the host and defense mechanisms of the immune system (Vizcaíno & Cloeckaert, 2012). Structurally, it is an asymmetrical lipid bilayer composed of lipopolysaccharide (LPS) and other haptenic polysaccharides, such as hapten native (NH), proteins and phospholipids (PL), with the LPS molecules and PL located in the outer and inner leaflet, respectively (**Figure 1.3**). Most of Gram-negative bacteria share this basic structure with OM molecules bearing pathogen-associated molecular patterns (PAMPs) easily recognized by innate immunity. However, brucellae OM has an atypical composition giving distinctive traits to the bacterium when compared with other Proteobacteria (Vizcaíno & Cloeckaert, 2012), and it has been proposed that some of its OM molecules, such as LPS, lipoproteins, flagellin and ornithine lipids, display a reduced PAMP (Gil-Ramírez *et al.*, 2014; Palacios-Chaves *et al.*, 2011; Barquero-Calvo *et al.*, 2007).

The different surface molecules are involved in various important processes, such as bacterial growth, sensing of and protection from environmental stresses, adhesion, and invasion of host cells, signaling, and interaction with the immune system (Bierne & Cossart, 2007). Likewise, a comprehensive characterization of the OM repertoire is required to better understand the factors that contribute to the success of a bacterial pathogen in colonizing different environmental niches and its mammalian host. Smooth brucellae such as *B. abortus, B. melitensis* and *B. suis* have OMs that are unusually resistant to the disrupting action of bactericidal peptides and complement (Palacios-Chaves *et al.*, 2011). It is believed that this stability is related to several specific characteristics, namely: (1) brucellae OMs contain large amounts of phosphatidylcholine (PC) and blockage of the synthesis of PC with the subsequent replacement by phosphatidylethanolamine (PE, which is the major PL in Gram-negative bacteria) generates attenuation, suggesting that PC is essential for brucellae virulence (Conde-Álvarez *et al.*, 2006); (2) the presence of positively charged ornithine lipids (OL), locate in the

outer leaflet of the OM (Moriyón & López-Goñi, 1998), although recent studies indicate that OL have become dispensable in the existing brucellae. This is consistent within the trend observed in α-Proteobacteria animal pathogens to reduce and eventually eliminate the envelope components susceptible of recognition by innate immunity (Palacios-Chaves *et al.*, 2011); (3) the strong association of some outer membrane proteins (OMPs) to LPS and PG (Moriyón & López-Goñi, 1998), and (4) the presence of very long chain fatty acids (VLCFAs) in the lipid A of LPS that have the potential to span the OM displaying the terminal hydroxyl group in the periplasmic space, probably favoring a strong anchorage of the LPS and the integrity of the OM (Vizcaíno & Cloeckaert, 2012; Barquero-Calvo *et al.*, 2009).

In addition, it was confirmed that the *B. abortus* LPS core has a branched structure. Based on the observation that the mutation of glycosyltransferase WadC, results in a lipopolysaccharide that, while keeping the O-polysaccharide essential for optimal protection, shows a truncated core that is more readily recognized by elements of innate immunity, it was proposed that the *Brucella* LPS core branch is a virulence-related structure accounting in part for the stealthy behavior of these bacteria (Gil-Ramírez *et al.*, 2014; Kubler-Kielb & Vinogradov, 2013; Conde-Álvarez *et al.*, 2012). The characteristic surface properties make brucellae furtive pathogens and more resistant to several host defense compounds, contributing to the intracellular survival and its ability to establish chronic infection (Ruiz-Ranwez *et al.*, 2015).

The brucellae LPS possess unusual immunological properties such as low toxicity, high resistance to macrophage degradation and protection against immune responses, being a major virulence factor in *Brucella* (Lapaque *et al.*, 2005). Since LPS is the most relevant antigen during infection and vaccination, LPS and LPS-related molecules are extensively used in immunological studies and in the diagnosis of brucellosis (Moreno & Moriyón, 2006; Cardozo *et al.*, 2006; Aragón *et al.*, 1996). Among Gram-negative bacteria, the genus *Brucella* is the unique in which some species express the smooth (S)-type LPS (*B. abortus*, *B. melitensis*, *B. suis*, *B. microti*, *B. neotomae*, *B. ceti*, *B. pinnipedialis, B. inopinata*, *B. papionis* and *B. vulpis*) and others have naturally rough (R)-type LPS (*B. canis* and *B. ovis*). The S-LPS and R-LPS differ mostly in the most external LPS moiety (the O-polysaccharide), which is not synthetized in rough *Brucella* species and is also missing in the surface of the rough mutants of the smooth species that appear spontaneously during culture dissociation (Moreno & Moriyón, 2006). It is known that the production of a complete LPS is necessary for virulence of smooth *Brucella* strains (González *et al.*, 2008; Rittig *et al.*, 2003). However, the rough species *B. canis* and *B. ovis* are virulent in their respective preferred natural hosts and in animal models, and it was shown that rough brucellae attract and infect monocytes more effectively than smooth brucellae although only S-LPS phenotypes establish a specific host cell compartment permitting successful parasitism (Rittig *et al.*, 2003).

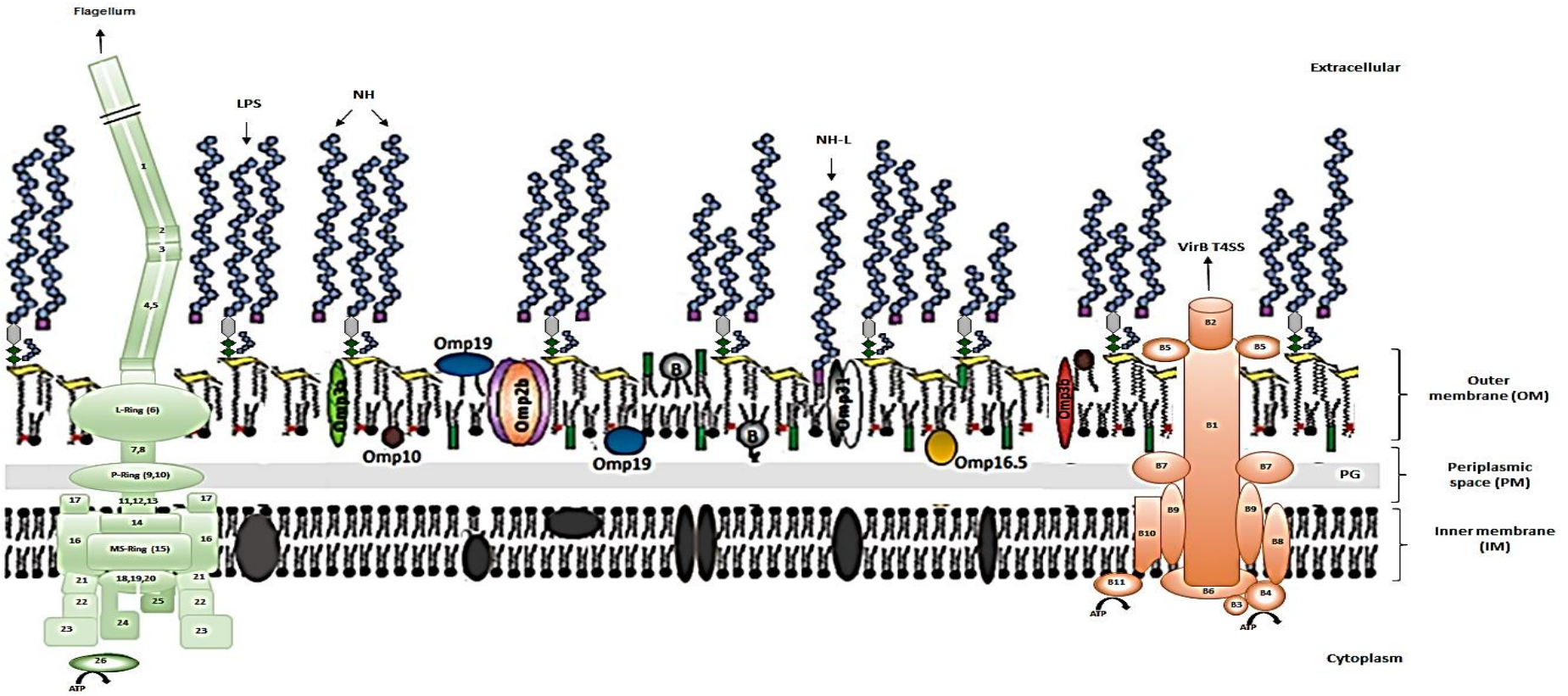


Figure 1.3. Schematic representation of Brucellae cellular membrane. Adapted and modified from Moreno & Moriyón (2006), Fabienne *et al.* (2008), Ferooz and Letesson (2010), Ferooz *et al.* (2011), Vizcaíno and Cloeckaert (2012), Gíl-Ramírez *et al.* (2014), Sankarasubramanian *et al.* (2016). The cellular membrane is constituted by the inner membrane (IM), the periplasmic space (PM) with the peptidoglycan (PG, light grey), and the outer membrane (OM). IM proteins are represented in grey; the OM contains the LPS (constituted by the lipid-A diaminoglucose disaccharide backbone represented by yellow trapezoids, the core oligosaccharide with its branch, and the O-polysaccharide chain), several outer membrane proteins (OMPs), ornithine lipids (green rectangles), phospholipids, mainly phosphatidylcholine (dark circles) and lipoproteins, namely the Braun's lipoprotein (B, grey circles) that can be linked to peptidoglycan or free in the OM. The purples squares mark the reducing ends of the native haptens (NH), an unknown sugar that in some cases may be linked to a lipid (NH-L). NH of various sizes are intertwined with the O-polysaccharides of the LPS in the OM, forming a dense layer. It is also shown the hypothetical model of the brucellae VirB Type IV secretion system (T4SS) and flagellum. B1 to B11 represent the 11 proteins that constitute the VirB T4SS. Numbers 1 to 26 symbolize proteins that compose the flagellum: 1.FliC (filament); 2.FlgL and 3.FlgK (Hook-filament junction); 4.FlgE and 5.FlgD (Hook); 6.FlgH, 7.FliL, 8.FlgG, 9. FlgL, 10.FlgA, 11.FlgB, 12.FlgC, 13.FlgF, 14.FliE and 15.FliF (Basal body); 16.MotA and 17.MotB (Motor); 18.FliP, 19.FliQ, 20.FliR, 21.FliG, 22.FliM, 23.FliN, 24.FlhA and 25.FlhB (Export apparatus); 26.FliI (ATP synthetase).

The most virulent species with higher zoonotic spectrum are those from domesticated animals, such as *B. melitensis*, *B. suis* (with the exception of biovar 2), *B. abortus* and *B. canis*; while those displaying lower pathogenicity and zoonotic potential are *B. ovis* and those from wildlife animals, like *B. suis* biovar 2 and biovar 5, *B. neotomae*, *B. microti*, *B. papionis*, *B. vulpis*, *B. pinnipedialis* and *B. ceti*. *Brucella* spp. are able to infect multiple hosts but some species are highly adapted to a single-host species, such *B. ovis* which is a pathogen for rams but does not infect other hosts (**Figure 1.4**) (Moreno, 2014; Olsen, 2014; Godfroid *et al.*, 2011; Tsolis *et al.*, 2009). In order to understand the pathomechanisms of infectious diseases with clinical significance in animals and humans we must first understand the biology of these agents, and their hosts, in order to unravel the interactions that occur at the host-agent interface. A comprehensive understanding of the basis of host specificity can provide insights into molecular pathogenesis, the evolution of pathogenic *Brucella* species, and the potential for these pathogens to cross the species barrier to infect new hosts.

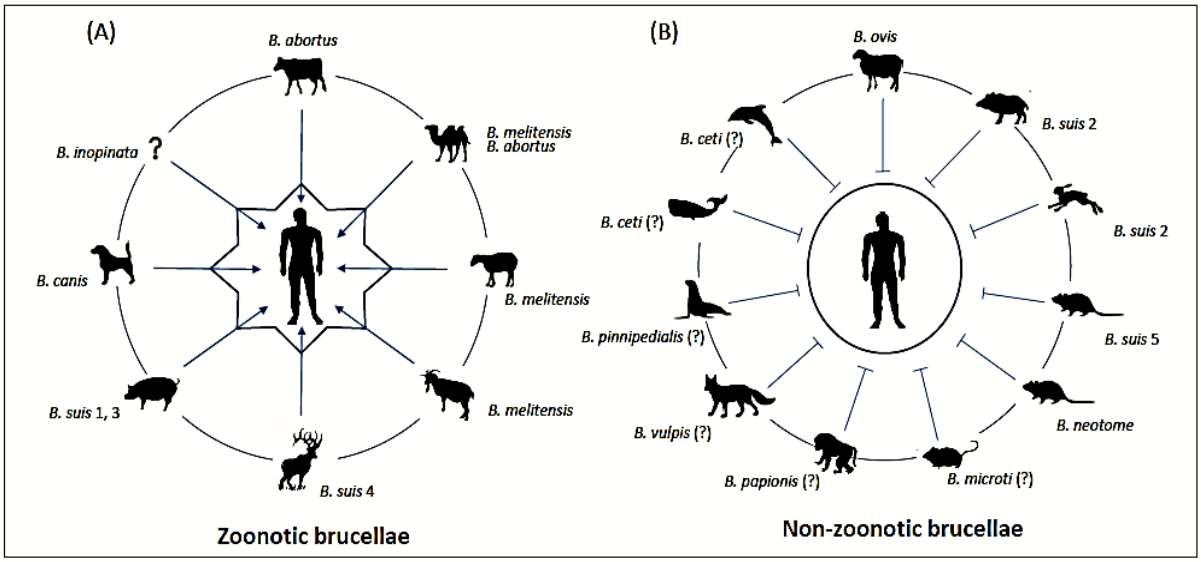


Figure 1.4. Zoonotic and non-zoonotic *Brucella* species, modified from Moreno, 2014. Hosts in clockwise direction: (A) cattle, camel, sheep, goat, reindeer, swine, dog, unknown; (B) rams, wild boar, hare, wild rodent, wild rodent, common vole, baboon, red fox, pinnipeds, porpoise, dolphin. Regarding *B. microti, B. papionis*, *B. vulpis, B. pinnipedialis* and *B. ceti*, pathogenicity for humans is still unknown.

One of the striking features that distinguish brucellae organisms from other pathogenic bacteria is that they do not display obvious virulence factors such as exotoxins, cytolysins, capsules, fimbria, plasmids, lysogenic phages, resistant forms, antigenic variation, endotoxic LPS or apoptosis inducers. Also unlike other pathogenic bacteria, brucellae virulence does not appear to be the result of relatively few virulence genes that can be transferred horizontally via plasmids, phages, or

assembled in pathogenicity islands. Instead, the true virulence elements of brucellae are those molecular determinants that allow to control their intracellular trafficking and adapt to the intracellular niche as well as on the extremely efficient adaptation to shield itself from the immune recognition and to manipulate key aspects of host cell physiology (apoptosis, vacuolar trafficking) (Ruiz-Ranwez *et al.*, 2015; Gomez *et al.*, 2013; von Bargen *et al.*, 2012; Lamontagne *et al.*, 2010; Rambow-Larsen *et al.*, 2009; Gorvel, 2008; Seleem *et al.*, 2008; Barquero-Calvo *et al.*, 2007; Moreno & Moriyón, 2006; Lapaque *et al.*, 2005; Gorvel and Moreno, 2002; Letesson *et al.*, 2002). However, in recent years, various virulence factors have been identified as essential for infection, including LPS (Cardozo *et al.*, 2006; Lapaque *et al.*, 2005; Ugalde *et al.*, 2000), β-cyclic glucan (Martirosyan *et al.*, 2012; Arellano-Reynoso *et al.*, 2005), BvrR/BvrS two component system (TCS) (Martín-Martín *et al.*, 2012; Lamontagne *et al.*, 2010; Viadas *et al.*, 2010; Guzman-Verri *et al.*, 2002), some Omps (Lim *et al.*, 2012; Vizcaíno and Cloeckaert, 2012), and the VirB Type IV secretion system (T4SS) (Seleem *et al.*, 2007; Celli *et al.*, 2003; Boschiroli *et al.*, 2002). Quorum sensing (QS) is also known to be involved in the regulation of brucellae virulence determinants mostly linked to the cell surface (T4SS, flagellum, Omps and exopolysaccharide), contributing to the adaptation of the metabolic network during the nutrient shift faced by brucellae all along its intracellular trafficking (Gorvel, 2014; Weeks *et al.*, 2010; Rambow-Larsen *et al.*, 2009; Letesson *et al.*, 2002).

In both humans and animals, brucellae first target the respiratory epithelium, the conjunctiva, and sexual organs. However, even nowadays, the mechanisms involved in brucellae entry into host cells still remain to be characterized (Gorvel, 2014). The ability of *Brucella* spp. to successfully survive and replicate within different host cells explains their pathogenicity. Extensive replication of *Brucella* spp. in placental trophoblasts is associated with abortion in their preferential animal hosts, and persistence in macrophages leads to chronic infections that are a hall mark of brucellosis in both natural animal hosts and humans (Kim, 2015; Gorvel, 2014; Grilló *et al.*, 2012; Roop *et al.*, 2009).

In vitro studies were used as models to understand adhesion, internalization, intracellular trafficking, survival, and replication of brucellae in susceptible hosts. After attachment to the epithelial cell surface receptors that contain sialic acid and sulfated residues, brucellae induces a zipper-like mechanism for internalization. Binding promotes activation of small GTPases that trigger a signaling cascade that reorganizes the actin cytoskeleton to induce a host cell membrane rearrangement along the surface of the pathogen that enhances invasion, and entry occurs within a few minutes after interaction (Rossetti *et al.*, 2012). *Brucella* spp. organisms are capable of colonizing macrophages, monocytes, and dendritic cells as well as trophoblasts, fibroblasts, endothelial cells, and epithelial cells (Gorvel, 2014; Hamer *et al.*, 2014; Martirosyan *et al.*, 2011; Starr *et al.*, 2008). Brucellae enters into host cells through lipid rafts (Barquero-Calve *et al.*, 2007; Porte *et al.*, 2003; Watarai *et al.*, 2002), and

its entry depends on the expression of BvrR/BvrS TCS (Manterola *et al.*, 2005; Guzman-Verri *et al.*, 2002). In fact, it has been shown that mutants lacking LPS O-chain do not use lipid rafts and are killed by the host cell suggesting that O-chain plays an important role in early events of host infection (Porte *et al.*, 2003). Brucellae survive and replicate inside nonprofessional phagocytic cells up to 72 hours *in vitro* and move across the epithelium *in vivo* by subverting the mucosal epithelial barrier function to facilitate brucellae migration. At the same time, this interaction initiates a minimal innate immune response with weak proinflammatory activity (Rossetti *et al.*, 2012; Barquero-Calvo *et al.*, 2007). Brucellae survival strategies have been elucidated from analysis of intracellular trafficking in either macrophage or epithelial cell models. Once internalized (**Figure 1.5**), *Brucella* spp. cells resides within the brucellae-containing vacuole (BCV), a modified phagosome in which the bacterium survives and ultimately proliferates (Kim *et al.*, 2015; Gomez *et al.*, 2013; Lee *et al.*, 2013; Starr *et al.*, 2008; Celli *et al.*, 2003; Chaves-Olarte *et al.*, 2002). The BCVs then fuse rapidly with the lysosome in a controlled manner, as suggested by the presence of the lysosomal markers, lysosomal-associated membrane protein (LAMP), and CD63, on the surface of bacteria (Starr *et al.*, 2008; Celli *et al.*, 2003; Pizarro-Cerdá *et al.*, 1998). In this transient stage, most of the contents of the BCVs are subjected to phagolysosomal degradation, and 90% of internalized brucellae cells are killed by the action of hydrolyzing enzymes (Celli *et al.*, 2003). However, the remaining 10% evade the host killing mechanisms through an unknown mechanism that probably involves the acidification of the BCVs, subsequent triggering of the virB operon and release of a large variety of effectors into the host cells' cytosol (Boschiroli *et al.*, 2002). The bacteria then traffic and arrive at the endoplasmic reticulum (ER). Within the ER, the bacteria survive and establish their replicative niche, and multiply to large numbers (Celli *et al.*, 2003). Recent studies have shown that autophagy-like vacuoles (*i.e.* autophagic “brucellae-replicating organelle”, aBCV) provide a replication-permissive compartment following the ER stage, that is essential for the completion of the intracellular lifecycle of brucellae and for cell-to-cell spreading (Starr *et al.*, 2008).

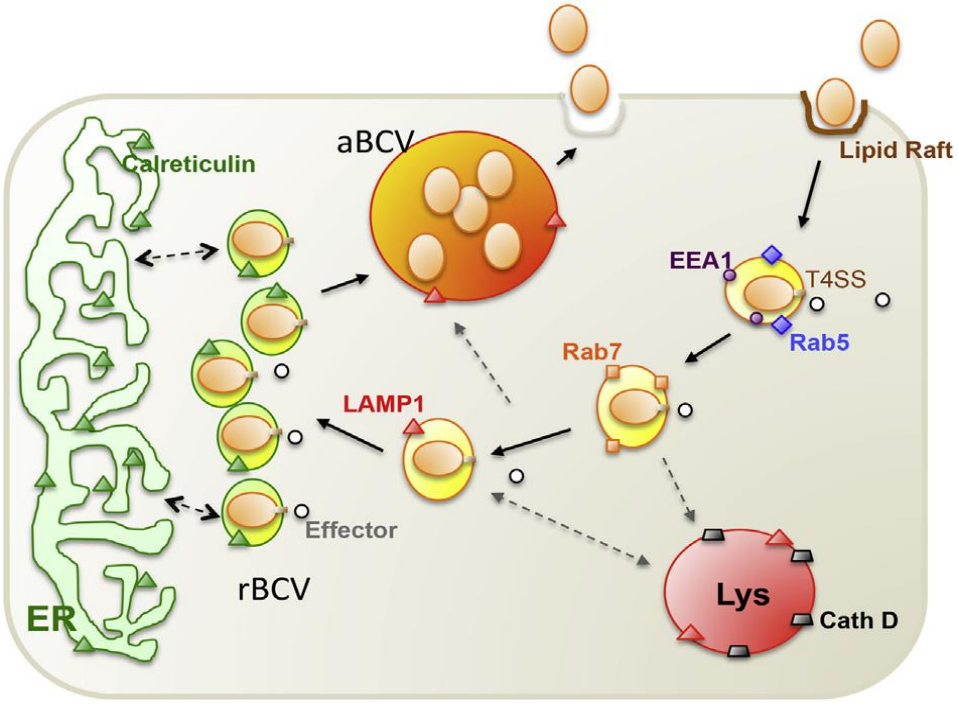


Figure 1.5. Working model of brucellae intracellular trafficking in macrophage cells, extracted from Kim, 2015. Plasma membrane-associated lipid rafts mediate the internalization of smooth brucellae into macrophage cells. As the BCV matures, it sequentially associates with markers for early (EEA1, purple circle; Rab5, blue diamond) and late (Rab7, orange square) endosomes. The biogenesis and trafficking of BCVs is regulated by bacterial effector proteins (white circles), which are secreted through the brucellae T4SS. BCVs that contain virulent organisms do not fuse with lysosomes (cathepsin D, gray trapezoid), although transient association with LAMP1-positive membranes (orange triangles) is observed. The pathogen replicates in tight rBCVs that are decorated with calreticulin (green triangle), a marker for the ER. At a later point after infection (48 to 72 hours), the pathogen is observed in LAMP1-positive aBCVs that also contain LAMP1. Finally, the pathogen is released from the cell through lytic or nonlytic (shown) mechanisms. aBCV, autophagic brucellae-containing vacuole; BCV, brucellae-containing vacuole; Beclin1, coiled-coil myosin-like BCL2-interacting protein; EEA1, early endosome antigen 1; ER, endoplasmic reticulum; LAMP1, lysosome-associated membrane protein 1; rBCV, replicative brucellae-containing vacuole; T4SS, type IV secretion system; ULK1, Unc-51-like kinase 1.

1.5. Comparative genomics as a tool to understand evolution in brucellae

The way in which we perceive the taxonomic relationships among different bacteria influences our understanding of their basic biological and ecological features (Moreno & Moriyón, 2002). Distinguishing individual bacterial lineages within a species, initially by phenotypic and subsequently by genotypic typing techniques, has been the cornerstone of infectious disease epidemiology, allowing the identification and tracking of the organisms responsible for infection and disease (Parkhill and Wren, 2011). During the past decade, the understanding of evolution at the genomic level has been

shaken to its core by many reports showing that genomes from closely related species can vary in terms of gene content (Williams *et al.*, 2010).

The genome from *B. melitensis* biovar 1 strain 16M was the first to be sequenced (DelVecchio *et al.*, 2002) followed by those from strains of *B. suis* biovar 1 strain 1330 (Paulsen *et al.*, 2002) and *B. abortus* biovar 1 strain 9-941 (Halling *et al.*, 2005). To date, genome sequences from more than 350 different *Brucella* strains, representing all species, have been published either as complete genomes or as draft assemblies. In 2009, the complete genome of *B. suis* biovar 2 (ATCC 23445) was sequenced and annotated, and, more recently, also the complete genomes of *B. suis* biovar 3 strain 686 (ATCC 23446), biovar 4 (ATCC 23447) and biovar 5 strain 513 (NCTC 11996) were released in NCBI. The sequences were very highly homologous, although regions of unique genetic material were also observed that could be involved in establishing the distinct host preferences and biological behavior of the different *Brucella* species (Lamontagne *et al.*, 2010). **Table 1.2** summarizes the general characteristics of the reference genomes for *B. abortus* (strain 9-941), *B. melitensis* (strain 16M), *B. suis* (strain 1330) and *B. suis* ATCC 23445 (strain Thomsen). The genomes of *B. abortus, B. melitensis* and *B. suis* biovar 1, like the majority of the *Brucella* species, consist of two circular chromosomes of 2.1 Mbp (Chr I) and 1.2 Mbp (Chr II), whereas *B. suis* biovar 2 has a smaller Chr I (1.9 Mbp) and a larger Chr II (1.4 Mbp). This last structure is also observed in the genome of the reference strain *B. suis* biovar 4 (strain 40). Moreover, a third structure was detected in *B. suis* biovar 3, strain 686, which have a single circular replicon of 3.3 Mbp. These three possible genomic structures appear to be the products of recombination events between the three *rrn loci* (**Figure 1.6**) (O'Callaghan and Whatmore, 2011; Jumas-Bilak *et al.*, 1998).

Genomic analysis indicates that the two chromosomes probably have distinct evolutionary origins. The origin of replication of Chr I is typical of bacterial circular chromosomes, while that of Chr II possesses a cluster of plasmid-like replication genes including a replication initiation protein RepC and partitioning proteins RepA and RepB, similar to plasmid replication genes from *Agrobacterium* Ti plasmids, and plasmids from other organisms including *Rhizobium* spp. (Paulsen *et al.*, 2002). Further, most of the essential genes for protein synthesis are located in Chr I while those encoding enzymes for sugar metabolism, protein regulators and membrane transport proteins for sugar, dipeptides and amino acids reside on Chr II (Halling *et al.*, 2004; Paulsen *et al.*, 2002).

Table 1.2. General characteristics of the reference genomes for *B. abortus* strain 9-941, *B. melitensis* strain 16M, *B. suis* strain 1330 and *B. suis* ATCC 23445 (strain Thomsen).

Genomic features [1]	*B. abortus* 9-941 (bv 1)		*B. melitensis* 16M (bv 1)		*B. suis* 1330 (bv 1)		*B. suis* ATCC 23445 (bv 2)	
	Chr I	Chr II	Chr I	Chr II	Chr I	Chr II	Chr I	Chr II
NCBI RefSeq	NC_006932	NC_006933	NC_003317	NC_003318	NC_004310	NC_004311	NC_010169	NC_010169
Length (bp)	2,124,241	1,162,204	2,117,144	1,177,787	2,107,792	1,207,381	1,923,763	1,400,844
Topology	Circular	Circular	Circular	Circular	Circular	Circular	Circular	Circular
% GC	57.2	57.3	57.2	57.3	57.2	57.3	57.2	57.3
Protein-coding gene	2,034	1,084	1,996	1,089	1,934	1,087	1,773	1,247
Pseudogenes	15	17	58	55	30	27	36	51
rRNA	6	3	6	3	6	3	3	6
tRNA	41	14	40	14	41	14	35	19

[1] Data obtained from NCBI; bv, biovar

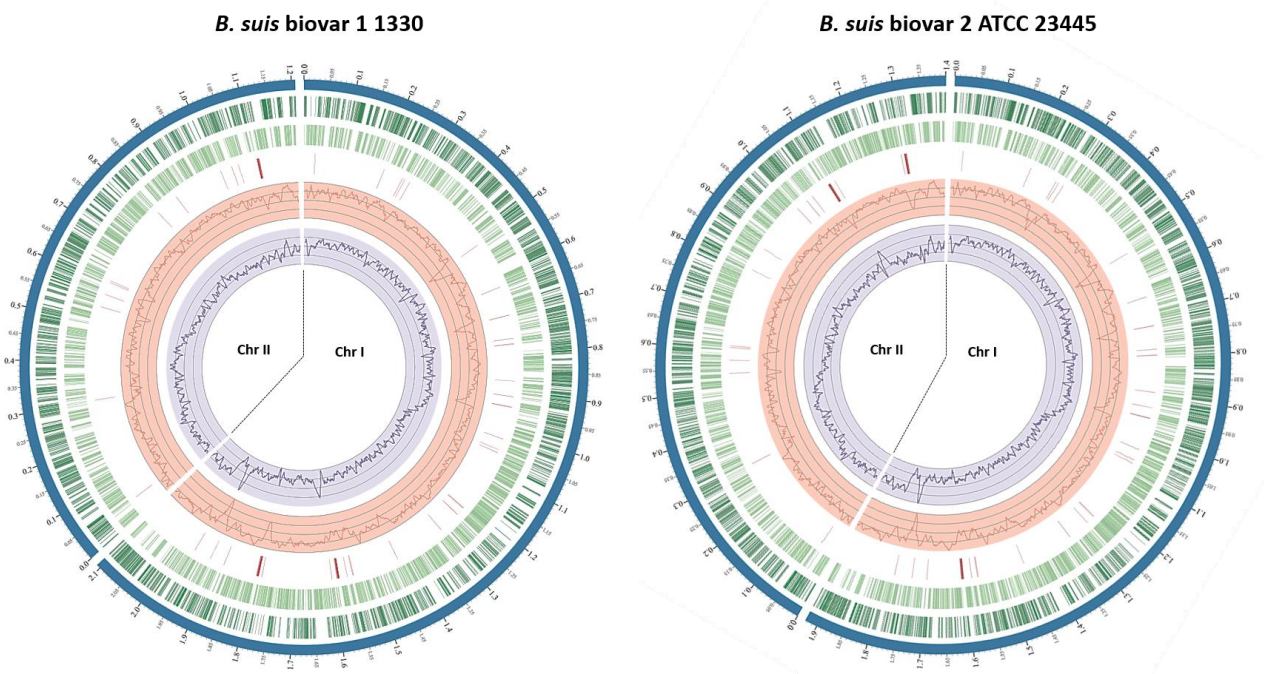


Figure 1.6. Graphical circular map of the genome for *B. suis* 1330 and *B. suis* ATCC 23445. From the outside to the center: Chromosomes I and II (blue), putative genes in forward (dark green), putative genes in reverse (green), tRNA (red), rRNA (dark red), GC plot (orange) and GC skew (purple). Circular graphics were obtained from the Pathosystems Resource Integration Center (PATRIC) website (www.patricbrc.org).

The similarity of the genomic sequences among *B. melitensis*, *B. suis*, *B. abortus*, *B. canis*, and *B. ovis* is evident by direct comparisons of their genomes showing more differences in Chr II than Chr I (Wattam *et al.*, 2009; Foster *et al.*, 2005; Halling *et al.*, 2004, 2005; Paulsen *et al.*, 2002). The major difference in Chr I was previously identified by physical mapping in *B. suis* biovar 2 strain Thomsen (ATCC 23445), where a 210 kb segment of Chr I has been translocated to Chr II (Jumas-Bilak *et al.*, 1998; Wattam *et al.*, 2009). For the somewhat more variable Chr II, more internal rearrangements, including one large chromosomal inversion of 640 kb in *B. abortus*, has been described (**Figure 1.7**) (Michaux-Charachon *et al.*, 1997; Halling *et al.*, 2005; Wattam *et al.*, 2009).

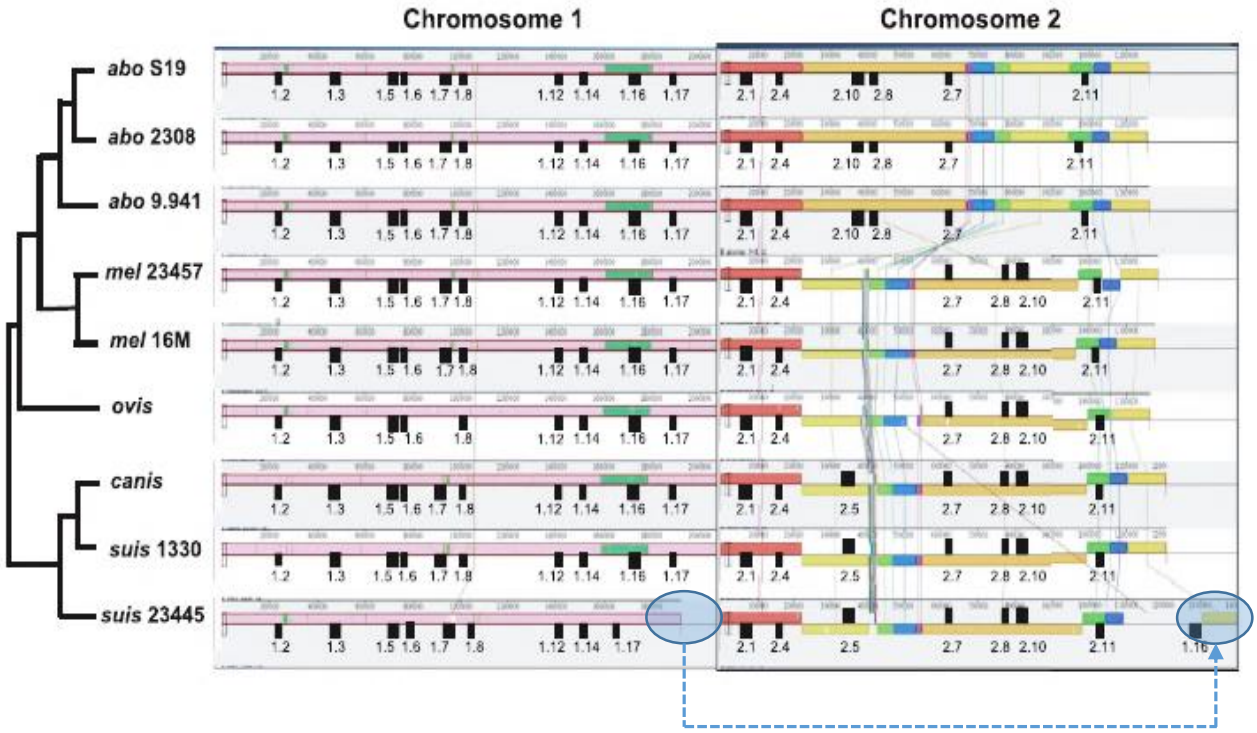


Figure 1.7. Mauve alignment of both chromosomes from nine complete *Brucella* genomes. A phylogenetic map of the strains, derived from the maximum likelihood tree and based on a concatenated alignment of 2,246 protein families (topology only, not branch lengths) is shown on the left side: abo, *B. abortus*; mel, *B. melitensis*; ovis, *B. ovis*; canis, *B. canis*; suis, *B. suis*. Shared anomalous regions (SARs) of interest are noted by filled boxes, with the names of those regions directly below them. The translocation event in *B. suis* ATCC 23445 is highlighted. Adapted from Wattam *et al.*, 2009.

The comparison of *B. suis* 1330 and *B. melitensis* 16M genomes defined a finite set of differences that could be responsible for the differences in virulence and host preference between these organisms and indicated that phage have played a significant role in their divergence. The sequencing of the entire genomes of these strains has revealed the existence of 33 DNA regions greater

than 100 bp specific to each strain. The discovery of *B. melitensis* 16M (11 regions) or *B. suis* 1330 (22 regions) specific DNA regions suggests that other specific DNA regions may also be present or absent in other *Brucellae*. Moreover, based on the whole genome alignments, only 42 *B. suis* genes were identified that are completely absent in the other genome (Paulsen *et al.* 2002).

Ratushna *et al.* (2006), carrying out a three-way genome comparison of *B. suis*, *B. melitensis* and *B. abortus* sequences, at both the nucleotide and predicted coding sequence (CDS) levels, found that the majority (>90%) of annotated genes were found to share 98–100% sequence identity and that most of differentiating genes identified are in large regions with functional assignments in existing annotation. Fewer than 100 genes were identified as present in only one or two of the three genomes, with an additional group of close to 100 genes having significant deletions in one or two of the genomes relative to the others. According to these authors, computational and experimental analysis identified and confirmed a set of 22 ORFs to be present in *B. suis* 1330, but not in *B. melitensis* 16 M or *B. abortus* 9-941, and another 22 ORFs found in both *B. suis* 1330 and *B. abortus* 9-941, but not in *B. melitensis* 16 M. Only *B. suis* was found to have a significant number of unique genes. Among these *B. suis* unique genes, two ABC transporters permeases and a cluster of transfer genes (*tra*/*trb*) and potentially significant to secretion (island S2) were identified (Ratushna *et al.*, 2006; Paulsen *et al.*, 2002).

Once limited genome diversity exists among different *Brucella* species, the analysis of full genome sequences of the different species (and biovars) is of crucial importance since it will enhance the knowledge on the biochemical pathways and will open the road for the identification of virulence factors. Because the genomes of *Brucella* species are highly conserved, host preference and virulence difference must stem from the limited genome diversity.

1.6. Objectives, research strategy and thesis organization

The work developed throughout this thesis aimed to improve our knowledge on *B. suis* biovar 2, the causative agent of swine brucellosis in Europe, and further unveil specific genomic variations associated with strains of specific geographic origin, namely, with Iberian strains. Further sequencing of the genomes of Iberian field strains seemed thus crucial to disclose the structure and distinctive features of *B. suis* pan-genome, as well as its evolutive history, host preference and geographic partitioning.

Although infection due to *B. suis* biovar 2 is recognized, few information is available about the strains circulating in swine and wild boar populations across Portugal, as well as on their genetic

diversity relatively to strains circulating in Spain and other European countries. Likewise, the first aim of this study was to develop a new selective medium in order to increase the sensitivity of the bacteriological diagnosis and consequently to enlarge the number of *B. suis* isolates in INIAV *Brucella* spp. culture collection. Subsequently, all isolates were identified and characterized using a polyphasic approach, in order to extend our understanding on the prevalence and molecular epidemiology of *B. suis* biovar 2 infection.

In accordance with other authors, the former studies showed that *B. suis* biovar 2 isolates were grouped in closely related clusters according to their geographic origins, distinguishing two biovar 2 clonal lineages in Iberian Peninsula. The third objective of this thesis was to evaluate the usefulness of optical mapping technology as an epidemiological tool. For that, the ordered restriction fragment maps (optical maps) of five *B. suis* biovar 2 strains representative of the two lineages circulating in Iberian Peninsula were constructed, with the purpose of disclosing genomic variations among them and assessing the universality of these markers, as well as the usefulness of this technology as an epidemiological tool in a large set of field strains.

To achieve a better understanding of the mechanisms of evolution and specialization of Iberian lineages, the last aim of this thesis was the full comparative genomic analysis of *B. suis* biovar 2 with other *Brucella* species to disclose the genomic and structural differences between Iberian and Central-European clonal lineages and further discuss the potential factors that favour the evolution of host specialization. The complete genomes of the five biovar 2 strains used for construction of optical maps were sequenced using the next-generation sequencing platform from Illumina (HiSeq 2000 system version 1.9), with a paired-end protocol. The full genome sequences were obtained using a combination of Illumina, Sanger and optical mapping technologies. The following comparative genomic analysis was performed with the intention of disclose common organizational and structural features, assess intra- and inter-biovar genomic diversity, and identify candidate-virulence genes and polymorphic regions. Polymorphic regions with marker potential (*e.g.* exclusive of *B. suis* Iberian clonal lineage) were further analysed to search for consensus regions suitable for the subsequent design of specific amplification primers. With this approach, it was intended to identify genetic variations that were consistently found associated with strains of a specific geographic origin. The genomic regions with wide distribution in field samples may also be elected as molecular markers with potential diagnostic and/or epidemiological value.

At the end of this PhD project, it was expected to have identified some specific genomic variations associated with strains of specific geographic origin, namely, with Iberian strains of *B. suis* biovar 2. However, other avenues can be explored and additional outcomes can be achieved, depending both on the obtained data and the temporal feasibility.

The wet-lab work was mostly developed at the Bacteriology and Mycology Laboratory from the Strategic Unity of Production and Animal Health from the Instituto Nacional de Investigação Agrária e Veterinária (INIAV, I.P.), with the cooperation of the Research Group on Microbiology & Biotechnology (M&B), from Biosystems and Integrative Sciences Institute (BioISI), Faculdade de Ciências, Universidade de Lisboa, Portugal.

Thesis organization

This thesis is organized into five chapters. In this **Chapter 1**, a general introduction to the subject of the thesis was presented, including the history and current status of knowledge, the objectives and the research strategy of the work developed in this thesis. **Chapter 2** is divided in two complementary studies. The first part (**subchapter 2.1**) includes the development and evaluation of a selective medium for the primary isolation of *B. suis*, LNIV-M, that has been developed and compared to the currently used selective media in the bacteriological diagnosis of brucellosis. This work allowed to follow up a bacteriological study which included the isolation of *B. suis* strains from wild boars and swine from different regions of Portugal, enhancing the number of isolates in the INIAV *Brucella* culture collection. In the second part (**subchapter 2.2**) is presented the molecular characterization of *B. suis* field strains using PCR-RFLP analysis for *omp2a, omp2b* and *omp31* genes and *Brucella* MLVA-16 assay, that endorsed the observation that two specific *B. suis* biovar 2 clonal lineages were circulating in Portugal and Spain, the Iberian and the Central-European clonal lineages. In **Chapter 3** it is described the application of optical maps to compare closely-related *B. suis* biovar 2 strains. Optical maps of five *B. suis* biovar 2 strains belonging to the Iberian and Central-European clonal lineages, as well as the optical maps of the *B. suis* biovar 2 reference strain, were produced allowing the identification of chromosomal rearrangements and insertions/deletions events, permitting to distinguish between clonal lineages and strains at the individual level. **Chapter 4** is divided in two separate works. In **subchapter 4.1** is presented the complete and annotated genome sequences of the five abovementioned *B. suis* biovar 2 strains from the two circulating clonal lineages, which have been released in GenBank. In **subchapter 4.2** a full genomic comparative analysis was performed, encompassing the genomes of the five *B. suis* biovar 2 strains belonging to the main circulating clonal lineages in Iberian Peninsula and the publicly available *Brucella* spp. genomes. Lastly, **Chapter 5** summarizes all major contributions from this PhD project and suggests directions for future research.

References

1. **Allen, A., Breadon, E., Byrne, A., Mallon, T., Skuce, R., Groussaud, P., Dainty, A., Graham, J., Jones K., Pollock, L. & Whatmore, A. (2015).** Molecular Epidemiology of *Brucella abortus* in Northern Ireland-1991 to 2012. PLoS One, 10(9): e0136721. doi: 10.1371/journal.pone.0136721.
2. **Al Dahouk, S., Whatmore, A.M., Hofer, E., Tomaso, H., Vergnaud, G., Le Flèche, P., Cloeckaert, A., Koylass, M.S., Whatmore, A.M., Nöckler, K. & Scholz, H.C. (2012).** Intraspecies biodiversity of the genetically homologous species *Brucella microti.* Applied Environmental Microbiology, 78(5):1534. doi: 10.1128/AEM.06351-11.
3. **Al Dahouk, S., Scholz, H.C., Tomaso, H., Bahn, P., Göllner, C., Karges, W., Appel, B., Hensel, A., Neubauer, H. & Nöckler, K. (2010)**. Differential phenotyping of *Brucella* species using a newly developed semi-automated metabolic system. BMC Microbiology, 10:269. doi:10.1186/1471-2180-10-269
4. **Alton, G.G., Jones, L.M., Angus, R.D. & Verger, J.M. (1988)**. Techniques for the Brucellosis Laboratory. Institute National de la Recherche Agronomique. Paris, France.
5. **Arellano-Reynoso, B., Lapaque, N., Salcedo, S., Briones, G., Ciocchini, A.E., Ugalde, R., Moreno, E., Moriyón, I. & Gorvel, J.P. (2005)**. Cyclic beta-1,2-glucan is a *Brucella* virulence factor required for intracellular survival. Nature Immunology, 6(6): 618-25.
6. **Aparício, E.D. (2013)**. Epidemiology of brucellosis in domestic animals caused by *Brucella melitensis*, *Brucella suis* and *Brucella abortus*. Revue Scientifique et Technique-Office International des Epizooties, 32: 53-60.
7. **Aragón, V., Díat, R., Moreno, E. & Moriyón, I. (1996)**. Characterization of *Brucella abortus* and *Brucella melitensis* native haptens as outer membrane 0-type polysaccharides independent from the smooth lipopolysaccharide. Journal of Bacteriology, 178: 1070-1079.
8. **Audic, S., Lescot, M., Claverie, J.M., Cloeckaert. A. & Zygmunt, M.S. (2011).** The genome sequence of *Brucella pinnipedialis* B2/94 sheds light on the evolutionary history of the genus *Brucella*. BMC Evolutionary Biology, 11:200. doi: 10.1186/1471-2148-11-200.
9. **Audic, S., Lescot, M., Claverie, J.M. & Scholz, H.C. (2009)**. *Brucella microti*: the genome sequence of an emerging pathogen. BMC Genomics, 10:352 doi:10.1186/1471-2164-10-352.
10. **Banai, M. & Corbel, M. (2010)**. Taxonomy of *Brucella*. The Open Veterinary Science Journal, 4: 85-101.
11. **Barquero-Calvo, E., Chaves-Olarte, E., Weiss, D.S., Guzman-Verri, C., Chacon-Díaz, C., Rucavado, A., Moriyón, I. & Moreno, E. (2007)**. *Brucella abortus* uses a stealthy strategy to avoid activation of the innate immune system during the onset of infection. PLoS One, 2: e631.
12. **Bergagna, S., Zoppi, S., Ferroglio, E., Gobetto, M., Dondo, A., Di Giannatale, E., Gennero, M.S. & Grattarola, C. (2009)**. Epidemiologic survey for *Brucella suis* biovar 2 in a wild boar (Sus scrofa) population in northwest Italy. Journal of Wildlife Diseases, 45(4): 1178-81.
13. **Bergey, D.H. & Holt, J.G. (1994)**. Bergey's Manual of Determinative Bacteriology. 9. Williams & Wilkins; Baltimore, MD, USA.
14. **Bierne, H. & Cossart, P. (2007)**. *Listeria monocytogenes* surface proteins: from genome predictions to function. Microbiology and Molecular Biology Reviews, 71(2): 377-97.
15. **Bohlin, J., Snipen, L., Cloeckaert, A., Lagesen, K., Ussery, D., Kristoffersen, A.B. & Godfroid J. (2010).** Genomic comparisons of *Brucella* spp. and closely related bacteria using base compositional and proteome based methods. BMC Evolutionary Biology, 10:249. doi: 10.1186/1471-2148-10-249.
16. **Boschiroli, M.L., Ouahrani-Bettache, S., Foulongne, V., Michaux-Charachon, S., Bourg, G., Allardet-Servent, A., Cazevieille, C., Lavigne, J.P., Liautard, J.P., Ramuz, M. & O'Callaghan, D. (2002)**. Type IV secretion and *Brucella* virulence. Veterinary Microbiology, 90(1-4): 341-8.
17. **Cardozo, P.G., Macedo, G.C., Azevedo, V. & Oliveira, S.C. (2006)**. *Brucella* spp noncanonical LPS: structure, biosynthesis, and interaction with host immune system. Microbial Cell Factories, 5:13 doi:10.1186/1475-2859-5-13.

18. **Carvalho, F.M., Souza, R.C., Barcellos, F.G., Hungria, M. & Vasconcelos, A.T.R. (2010)**. Genomic and evolutionary comparisons of diazotrophic and pathogenic bacteria of the order Rhizobiales. BMC Microbiology, 10: 37. doi: 10.1186/1471-2180-10-37.
19. **Celli, J., de Chastellier, C., Franchini, D.M., Pizarro-Cerda, J., Moreno, E. & Gorvel, J.P. (2003)**. *Brucella* evades macrophage killing via VirB-dependent sustained interactions with the endoplasmic reticulum. The Journal of Experimental Medicine, 198(4):545-56.
20. **Cloeckaert, A. & Vizcaino, N. (2004)**. DNA Polymorphism and Taxonomy of *Brucella* Species. *In Brucella*: Molecular and Cellular Biology. Chapter 1, pag. 1-24. Edited by Ignacio López-Goñi and Ignacio Moriyón.
21. **Chain, P.S., Comerci, D.J., Tolmasky, M.E., Larimer, F.W., Malfatti, S.A., Vergez, L.M., Aguero, F., Land, M.L., Ugalde, R.A. & Garcia E. (2005)**. Whole-genome analyses of speciation events in pathogenic Brucellae. Infection and Immunity, 73(12):8353-61.
22. **Chaves-Olarte, E., Guzmán-Verri, C., Méresse, S., Desjardins, M., Pizarro-Cerdá, J., Badilla, J., Gorvel, J.P. & Moreno, E. (2002)**. Activation of Rho and Rab GTPases dissociates *Brucella abortus* internalization from intracellular trafficking. Cell Microbiology, 4(10):663-76.
23. **Conde-Álvarez, R., Arce-Gorvel, V., Iriarte, M., Manček-Keber, M., Barquero-Calvo, E., Palacios-Chaves, L., Chacón-Díaz, C., Chaves-Olarte, E., Martirosyan, A., von Bargen, K., Grilló, M.J., Jerala, R., Brandenburg, K., Llobet, E., Bengoechea, J.A., Moreno, E., Moriyón, I. & Gorvel, J.P. (2012)**. The lipopolysaccharide core of *Brucella abortus* acts as a shield against innate immunity recognition. PLoS Pathogens, 8(5):e1002675. doi: 10.1371/journal.ppat.1002675.
24. **Cvetnic, Z., Mitak, M., Ocepek, M., Lojkic, M., Terzic, S., Jemersic, L., Humski, A., Habrun, B., Sostaric, B., Brstilo, M., Krt, B. & Garin-Bastuji, B. (2003)**. Wild boars (Sus scrofa) as reservoirs of *Brucella suis* biovar 2 in Croatia. Acta Veterinaria Hungarica, 51(4):465–473.
25. **De Ley, J., Mannheim, W., Segers, P., Lievens, A., Denijn, M., Vanhoucke, M. & Gillis, M. (1987**). Ribosomal Ribonucleic Acid Cistron Similarities and Taxonomic Neighborhood of *Brucella* and CDC Group Vd. International Journal of Systematic and Bacteriology, 37 (1): 35-42.
26. **DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., Reznik, G., Jablonski, L., Larsen, N, D'Souza, M., Bernal, A., Mazur, M., Goltsman, E., Selkov, E., Elzer, P.H., Hagius, S., O'Callaghan, D., Letesson, J.J, Haselkorn, R., Kyrpides, N. & Overbeek, R. (2002)**. The genome of the facultative intracellular pathogen *Brucella melitensis*. Proceedings of the National Academy of Sciences of the United States of America, 99(1): 443-448.
27. **Eisenberg, T., Rißе, K., Schauerte, N., Geiger, C., Blom, J. & Scholz, H.C. (2016)**. Isolation of a novel 'atypical' *Brucella* strain from a bluespotted ribbontail ray (*Taeniura lymma*). Antonie Van Leeuwenhoek, 110(2): 221-234. doi: 10.1007/s10482-016-0792-4.
28. **Eisenberg, T., Hamann, H.P., Kaim U., Schlez, K., Seeger, H., Schauerte, N., Melzer, F., Tomaso, H., Scholz, H.C., Koylass, M.S., Whatmore, A.M. & Zschock M. (2012)**. Isolation of potentially novel *Brucella* spp. from frogs. Applied and Environmental Microbiology, 78 (10), 3753–3755
29. **EFSA (2009)**. Scientific Opinion of the Panel on Animal Health and Welfare (AHAW) on a request from the Commission on porcine brucellosis (*Brucella suis*). The EFSA Journal, 1144-1112.
30. **Ferreira, A.C., Chambel, L., Tenreiro, T., Cardoso, R., Flor, L., Dias, I.T., Pacheco, T., Garin-Bastuji, B., Le Flèche, P., Vergnaud, G., Tenreiro, R. & de Sá, M.I. (2012)**. MLVA16 typing of Portuguese human and animal *Brucella melitensis* and *Brucella abortus* isolates. PLoS One. 7: e42514.
31. **Ficht, T. (2010)**. *Brucella* taxonomy and evolution. Future Microbiology, 5: 859–866.
32. **Foster, J.T., Price, L.B., Beckstrom-Sternberg, S.M., Pearson, T., Brown, W.D., Kiesling, D.M., Allen, *C.A.*, Liu, C.M., Beckstrom-Sternberg, J., Roberto, F.F. & Keim, P. (2012).** Genotyping of *Brucella* species using clade specific SNPs. BMC Microbiology, 12: 110.
33. **Foster, J.T., Beckstrom-Sternberg, S.M., Pearson, T., Beckstrom-Sternberg, J.S., Chain, P.S., Roberto, F.F., Hnath, J., Brettin, T. & Keim P. (2009)**. Whole-genome-based phylogeny and divergence of the genus *Brucella*. Journal of Bacteriology, 191(8):2864-70. doi: 10.1128/JB.01581-08.

34. **Foster, J.T., Okinaka, R.T., Svensson, R., Shaw, K., De, B.K., Robison, R.A., Probert, W.S., Kenefic , L.J., Brown, W.D. & Keim, P. (2008)**. Real-time PCR assays of single-nucleotide polymorphisms defining the major *Brucella* clades. Journal of Clinical Microbiology, 46(1):296-301.
35. **Foster, G., Osterman, B.S., Godfroid, J., Jacques, I. & Cloeckaert, A. (2007).** *Brucella ceti* sp. nov. and *Brucella pinnipedialis* sp. nov. for *Brucella* strains with cetaceans and seals as their preferred hosts. International Journal of Systematic and Evolutionary Microbiology, 57: 2688-2693.
36. **Galindo, R.C., Muñoz, P.M., de Miguel, M.J., Marin, C.M., Labairu, J., Revilla, M., Blasco, J.M., Gortazar, C. & de la Fuente, J. (2010)**. Gene expression changes in spleens of the wildlife reservoir species, Eurasian wild boar (*Sus scrofa*), naturally infected with *Brucella suis* biovar 2. Journal of Genetics and Genomics, 37(11): 725-36. doi: 10.1016/S1673-8527(09)60090-4.
37. **García-Yoldi, D., Le Fleche, P., Marín, C.M., De Miguel, M.J., Muñoz, P.M., Vergnaud, G. & López-Goñi, I. (2007)**. Assessment of genetic stability of *Brucella melitensis* Rev 1 vaccine strain by multiple-*locus* variable-number tandem repeat analysis. Vaccine, 25: 2858-62.
38. **García-Yoldi, D., Marín, C.M., De Miguel, M. J., Muñoz, P.M., Vizmanos, J.L. & López-Goñi, I. (2006)**. Multiplex PCR assay for the identification and differentiation of all *Brucella* species and the vaccine strains *Brucella abortus* S19 and RB51 and *Brucella melitensis* Rev1. Clinical Chemistry, 52:779–781.
39. **Gee, J.E., De, B.K., Levett, P.N., Whitney, A.M., Novak, R.T. & Popovic, T. (2004)**. Use of 16S rRNA gene sequencing for rapid confirmatory identification of *Brucella* isolates. Journal of Clinical Microbiology, 42(8):3649–3654.
40. **Gil-Ramírez, Y., Conde-Álvarez, R., Palacios-Chaves, L., Zúñiga-Ripa, A., Grilló, M.-J., Arce-Gorvel, V., Hanniffy, S., Moriyón, I. & Iriarte, M. (2014)**. The identification of wadB, a new glycosyltransferase gene, confirms the branched structure and the role in virulence of the lipopolysaccharide core of *Brucella abortus*. Microbial Pathogenesis, 73: 53–59.
41. **Godfroid, J., Garin-Bastuji, B., Saegerman, C. & Blasco, J.M. (2013)**. Brucellosis in terrestrial wildlife. Revue Scientifique et Technique-Office International des Epizooties., 32 (1): 27-42
42. **Godfroid, J., Scholz, H.C., Barbierd, T., Nicolas, C., Wattiau, P., Fretin, D., Whatmore, A.M., Cloeckaert, A., Blasco, J.M., Moriyon, I., Saegerman, C., Muma, J.B., Al Dahouk, S., Neubauer, H. & Letesson, J.J. (2011)**. Brucellosis at the animal/ ecosystem/ human interface at the beginning of the 21st century. Preventive Veterinary Medicine, 102: 118-31.
43. **Godfroid, J. & Käsbohrer, A.** (**2002)**. Brucellosis in the European Union and Norway at the turn of the twenty-first century. Veterinary Microbiology, 90:135-145.
44. **Gomez, G., Adams, L.G., Rice-Ficht, A. & Ficht, T.A. (2013)**. Host-*Brucella* interactions and the *Brucella* genome as tools for subunit antigen discovery and immunization against brucellosis. Frontiers in Cellular and Infection Microbiology, 3:17. doi: 10.3389/fcimb.2013.00017.
45. **González, D., Grilló, M.J., De Miguel, M., Ali, T., Arce-Gorvel, V., Delrue, R.M., Conde-Álvarez, R., López-Goñi, I., Iriarte, M., Marín, C.M., Weintraub, A., Widmalm, G., Zygmunt, M.S., Letesson, J.J., Gorvel, J.P., Blasco, J.M. & Moriyón I. (2008)**. Brucellosis vaccines: assessment of *Brucella melitensis* lipopolysaccharide rough mutants defective in core and O-polysaccharide synthesis and export. PLoS ONE, 3:e2760.
46. **Gorvel, J.-P. (2014)**. If you bring an alarm, we will destroy it, said *Brucella* to the host cell. Virulence, 5: 460-462, DOI: 10.4161/viru.29092.
47. **Gorvel, J.P. (2008)**. *Brucella*: a Mr "Hide" converted into Dr Jekyll. Microbes and Infection, 10(9): 1010-3. doi: 10.1016/j.micinf.2008.07.007.
48. **Gorvel, J.P. & Moreno, E. (2002)**. *Brucella* intracellular life: from invasion to intracellular replication. Veterinary Microbiology, 90(1-4):281-97.
49. **Grilló, M.J., Blasco, J.M., Gorvel, J.P., Moriyón, I. & Moreno, E. (2012)**. What have we learned from brucellosis in the mouse model? Veterinary Research, 43: 29. doi: 10.1186/1297-9716-43-29.
50. **Guzman-Verri, C., Manterola, L., Sola-Landa, A., Parra, A., Cloeckaert, A., Garin, J., Gorvel, J.P., Moriyon, I., Moreno, E. & López-Goñi, I. (2002)**. The two-component system BvrR/BvrS essential for *Brucella abortus* virulence regulates the expression of outer membrane proteins with counterparts in

members of the Rhizobiaceae. Proceedings of the National Academy of Sciences of the United States of America, 99:12375-80;

51. **Halling, S.M., Peterson-Burch, B.D., Bricker, B.J., Zuerner, R.L., Qing, Z., Li, L.L., Kapur, V., Alt, D.P. & Olsen, S.C. (2005)**. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. Journal of Bacteriology, 187 (8): 2715-2726.
52. **Hamer, I., Goffin, E., De Bolle, X., Letesson, J.J., Jadot, M. (2014)**. Replication of *Brucella abortus* and *Brucella melitensis* in fibroblasts does not require Atg5-dependent macroautophagy. BMC Microbiology, 14:223. doi: 10.1186/s12866-014-0223-5.
53. **Hammerl, J.A., Göllner, C., Al Dahouk, S., Nöckler, K., Reetz, J. & Hertwig, S. (2016).** Analysis of the First Temperate Broad Host Range *Brucella* phage (BiPBO1) Isolated from *B. inopinata*. Frontiers in Microbiology, 7:24. doi: 10.3389/fmicb.2016.00024.
54. **Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., O'Callaghan, D. and Ramuz, M. (1998).** Differences in chromosome number and genome rearrangements in the genus *Brucella*. Molecular Microbiology, 27(1): 99–106.
55. **Kim, S. (2015)**. The interaction between *Brucella* and the host cell in phagocytosis. *In* Updates on Brucellosis. Chapter 3, pag. 45-60. Edited by Manal Mohammad Baddour, ISBN 978-953-51-2211-1, Publisher: InTech.
56. **Kubler-Kielb, J. & Vinogradov, E. (2013)**. The study of the core part and non-repeating elements of the O-antigen of *Brucella* lipopolysaccharide. Carbohydrate Research, 366:33-7. doi: 10.1016/j.carres.2012.11.004.
57. **Lamontagne, J., Béland, M., Forest, A., Côté-Martin, A., Nassif, N., Tomaki, F., Moriyón, I., Moreno, E. & Paramithiotis, E. (2010)**. Proteomics-based confirmation of protein expression and correction of annotation errors in the *Brucella abortus* genome. BMC Genomics, 11: 300. doi: 10.1186/1471-2164-11-300.
58. **Lapaque, N., Moriyon, I., Moreno, E. & Gorvel, J.-P. (2005)**. *Brucella* lipopolysaccharide acts as a virulence factor. Current Opinion in Microbiology, 8: 60–66.
59. **Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoed, P., Nöckler, K., Neubauer, H., Guilloteau, L.A. & Vergnaud, G. (2006)**. Evaluation and selection of tandem repeat *loci* for a *Brucella* MLVA typing assay. BMC Microbiology, 6:9-22.
60. **Letesson, J.J., Lestrate, P., Delrue, R.M., Danese, I., Bellefontaine, F., Fretin, D., Taminiau, B., Tibor, A., Dricot, A., Deschamps, C., Haine, V., Leonard, S., Laurent, T., Mertens, P., Vandenhaute, J. & De Bolle, X. (2002)**. Fun stories about *Brucella*: the "furtive nasty bug". Veterinary Microbiology, 90(1-4): 317-28.
61. **Leuenberger, R., Boujon, P., Thur, B., Miserez, R., Garin-Bastuji, B., Rufenacht, J. & Stark, K.D. (2007)**. Prevalence of classical swine fever, Aujeszky's disease and brucellosis in a population of wild boar in Switzerland. Veterinary Record, 160: 362–368.
62. **Lim, A., Dimalanta, E. T., Potamousis, K. D., Yen, G., Apodoca, J., Tao, C., Lin, J., Qi, R., Skiadas, J., Ramanathan, A., Perna, N.T., Plunkett III, G., Burland, V., Mau, B., Hackett, J., Blattner, F.R., Anantharaman, T.S., Mishra, B. & Schwartz, D.C. (2001)**. Shotgun optical maps of the whole *Escherichia coli* O157:H7 genome. Genome Research, 11:1584–1593.
63. **López-Goñi, I., García-Yoldi, D., Marín, C.M., de Miguel, M.J., Barquero-Calvo, E., Guzmán-Verri, C., Albert, D. & Garin-Bastuji, B. (2011)**. New Bruce-ladder multiplex PCR ass ay for the biovar typing of *Brucella suis* and the discrimination of *Brucella suis* and *Brucella canis*. Veterinary Microbiology, 154 (1–2): 152–155.
64. **López-Goñi, I., García-Yoldi, D., Marín, C.M., de Miguel, M.J., Muñoz, P.M., Blasco, J.M., Jacques, I., Grayon, M., Cloeckaert, A., Ferreira, A.C., Cardoso, R., Corrêa de Sá, M.I., Walravens, K., Albert, D. & Garin-Bastuji, B. (2008)**. Evaluation of a multiplex PCR assay (Bruce-ladder) for molecular typing of all *Brucella* species, including the vaccine strains. Journal of Clinical Microbiology, 46(10):3484-7.

65. **Manterola, L., Moriyón, I., Moreno, E., Sola-Landa, A., Weiss, D.S., Koch, M.H., Howe, J., Brandenburg, K. & López-Goñi, I. (2005)**. The lipopolysaccharide of *Brucella abortus* BvrS/BvrR mutants contains lipid A modifications and has higher affinity for bactericidal cationic peptides. Journal of Bacteriology, 187: 5631–5639.
66. **Martirosyan, A., Moreno, E. & Gorvel, J.-P. (2011)**. An evolutionary strategy for a stealthy intracellular *Brucella* pathogen. Immunological Reviews, 240: 211–234.
67. **Martín-Martín, A.I., Sancho, P., Tejedor, C., Fernández-Lago, L. & Vizcaíno, N. (2011)**. Differences in the outer membrane-related properties of the six classical *Brucella* species. The Veterinary Journal, 189: 103–105.
68. **Mayer-Scholl, A., Draeger, A., Gollner, C., Scholz, H.C. & Nockler K. (2010)**. Advancement of a multiplex PCR for the differentiation of all currently described *Brucella* species. Journal of Microbiological Methods, 80(1): 112–114.
69. **Michaux-Charachon, S., Bourg, G., Jumas-Bilak, E., Guigue-Talet, P., Allardet-Servent, A., O'Callaghan, D. & Ramuz, M. (1997)**. Genome structure and phylogeny in the genus *Brucella*. Journal of Bacteriology, 79: 3244-3249.
70. **Moreno, E. (2014)**. Retrospective and prospective perspectives on zoonotic brucellosis. Frontiers in Microbiology, 5: 213. doi: 10.3389/fmicb.2014.00213.
71. **Moreno, E. & Moriyón, I. (2006)**. The Genus *Brucella*. *In* Falkow S., *et al.*, editors. (ed.), The prokaryotes, vol. 5, p.315-456 Springer-Verlag, New York, NY.
72. **Moreno, E. & Moriyón, I. (2002)**. *Brucella melitensis*: a nasty bug with hidden credentials for virulence. Proceedings of the National Academy of Sciences of the United States of America, 99(1): 1-3. www.pnas.org/cgi/doi/10.1073/pnas.022622699.
73. **Moriyón, I. & López-Goñi, I. (1998).** Structure and properties of the outer membranes of *Brucella abortus* and *Brucella melitensis*. Internatl. Microbiol., 1: 19-26.
74. **Mühldorfer, K., Wibbelt, G., Szentiks, *C.A.*, Fischer, D., Scholz, H.C., Zschöck, M. & Eisenberg, T. (2016)**. The role of 'atypical' *Brucella* in amphibians: are we facing novel emerging pathogens? Journal of Applied Microbiology, doi:10.1111/jam.13326.
75. **Muñoz, P.M., Boadella, M., Arnal, M., de Miguel, M.J., Revilla, M., Martinez, D., Vicente, J., Acevedo, P., Oleaga, A., Ruiz-Fons, F., Marín, C.M., Prieto, J.M., de la Fuente, J., Barral, M., Barberán, M., de Luco, D.F., Blasco, J.M. & Gortázar, C. (2010)**. Spatial distribution and risk factors of Brucellosis in Iberian wild ungulates. BMC Infectious Diseases, 10:46.
76. **Musser, J.M., Schwartz, A.L., Srinath, I. & Waldrup, K.A. (2013)**. Use of serology and bacterial culture to determine prevalence of *Brucella* spp. in feral Swine (*Sus scrofa*) in proximity to a beef cattle herd positive for *Brucella suis* and *Brucella abortus*. Journal of Wildlife Diseases, 49:215–220.
77. **O'Callaghan, D. & Whatmore, A. (2011)**. *Brucella* genomics as we enter the multi-genome era. Briefings in Functional Genomics, 10(6): 334-341. doi: 10.1093/bfgp/elr026.
78. **OIE Terrestrial Manual (2016)**. Brucellosis (*Brucella abortus, B. melitensis* and *B. suis*) (Infection with *Brucella abortus, B. melitensis* and *B. suis*). Manual of Diagnostic Tests and Vaccines for Terrestrial Animals, chapter 2.1.4.
79. **Olsen, S.C. & Palmer, M.V. (2014)**. Advancement of knowledge of *Brucella* over the past 50 years. Veterinary Pathology, 1-14, doi:10.1177/0300985814540545.
80. **Osterman, B. & Moriyón, I. (2006)**. International Committee on Systematics of Prokaryotes: Subcommittee on the taxonomy of *Brucella*. International Journal of Systematic and Evolutionary Microbiology, 56:1173–1175.
81. **Palacios-Chaves, L., Conde-Álvarez, R., Gil-Ramírez, Y., Zúñiga-Ripa, A., Barquero-Calvo, E., Chacón-Díaz, C., Chaves-Olarte, E., Arce-Gorvel, V., Gorvel, J.P., Moreno, E., de Miguel, M.J., Grilló, M.J., Moriyón, I., Iriarte, M. (2011)**. *Brucella abortus* ornithine lipids are dispensable outer membrane components devoid of a marked pathogen-associated molecular pattern. PLoS One, 6(1): e16030. doi: 10.1371/journal.pone.0016030.

82. **Parkhill, J. & Wren, B.W. (2011)**. Bacterial epidemiology and biology - lessons from genome sequencing. Genome Biology, 12:230.
83. **Paulsen, I.T., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C,, Ayodeji, B., Kraul, M., Shetty, J., Malek, J., Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M. & Fraser, C.M. (2002)**. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proceedings of the National Academy of Sciences of the United States of America, 99 (20): 13148-13153.
84. **Plotkin, S.A. (2009)**. Vaccines: the Fourth Century. Clinical and Vaccine Immunology, 16:1709.
85. **Porte, F., Naroeni, A., Ouahrani-Bettache, S., Liautard, J.P. (2003)**. Role of the *Brucella suis* lipopolysaccharide O antigen in phagosomal genesis and in inhibition of phagosome–lysosome fusion in murine macrophages. Infection and Immunity, 71: 1481–1490.
86. **Rajashekara, G., Glasner, J.D., Glover, D.A. & Splitter, G.A. (2004).** Comparative whole-Genome hybridization reveals genomic islands in *Brucella* species. Journal of Bacteriology, 186(15): 5040–5051. doi: 10.1128/JB.186.15.5040–5051.2004.
87. **Rambow-Larsen, A.A., Petersen, E.M., Gourley, C.R. & Splitter, G.A. (2009)**. *Brucella* regulators: self-control in a hostile environment. Trends in Microbiology, 17(8):371-7. doi: 10.1016/j.tim.2009.05.006.
88. **Ratushna, V., Sturgill, D., Ramamoorthy, S., Reichow, S., He, Y., Lathigra, R., Sriranganathan, N., Halling, S., Boyle, S. & Gibas, C. (2006)**. Molecular targets for rapid identification of *Brucella* spp. BMC Microbiology, 6: 13-32.
89. **Rittig, M.G., Kaufmann, A., Robins, A., Shaw, B., Sprenger, H., Gemsa, D., Foulongne, V., Rouot, B. & Dornand, J. (2003)**. Smooth and rough lipopolysaccharide phenotypes of *Brucella* induce different intracellular trafficking and cytokine/chemokine release in human monocytes. Journal of Leukocyte Biology, 74(6):1045-55.
90. **Roop, R.M., Gaines, J.M., Anderson, E.S., Caswell, C.C. & Martin, D.W. (2009)**. Survival of the fittest: how *Brucella* strains adapt to their intracellular niche in the host. Medical Microbiology and Immunology, 198(4): 221-38. doi: 10.1007/s00430-009-0123-8.
91. **Rossetti, *C.A.*, Drake, K.L. & Adams, L.G. (2012)**. Transcriptome analysis of HeLa cells response to *Brucella melitensis* infection: a molecular approach to understand the role of the mucosal epithelium in the onset of the *Brucella* pathogenesis. Microbes and Infection, 14:756e767.
92. **Scholz, H.C., Revilla-Fernández, S., Al Dahouk, S., Hammerl, J.A., Zygmunt, M.S., Cloeckaert, A., Koylass, M., Whatmore, A.M., Blom, J., Vergnaud, G., Witte, A., Aistleitner, K. & Hofer E. (2016)**. *Brucella vulpis* sp. nov., isolated from mandibular lymph nodes of red foxes (*Vulpes vulpes*). Int International Journal of Systematic and Evolutionary Microbiology, 66(5):2090-8.
93. **Scholz, H.C. & Vergnaud, G. (2013)**. Molecular characterization of *Brucella* species. Revue Scientifique et Technique-Office International des Epizooties, 32(1):149-62.
94. **Scholz, H.C., Kämpfer, P. & Cloeckaert, A. (2012)**. *Brucella*: relationship to other Alphaproteobacteria, current taxonomy and the emergence of new species. In *Brucella* – Molecular Microbiology and Genomics, chapter 1, pg. 1-12. Edited by I. López-Goñi and D. O`Callaghan, Caister Academic Press, Norfolk, UK.
95. **Scholz, H.C., Nöckler, K., Göllner, C., Bahn, P., Vergnaud, G., Tomaso, H., Al Dahouk, S., Kämpfer, P., Cloeckaert, A., Maquart, M., Zygmunt, M.S., Whatmore, A.M., Pfeffer, M., Huber, B., Busse, H.J. & De BK. (2010)**. *Brucella* inopinata sp. nov., isolated from a breast implant infection. Int International Journal of Systematic and Evolutionary Microbiology, 60: 801-808.
96. **Scholz, H.C., Hubalek, Z., Sedlacek, I., Vergnaud, G., Tomaso, H., Al Dahouk, S., Melzer, F., Kämpfer, P., Neubauer, H., Cloeckaert, A., Maquart, M., Zygmunt, M.S., Whatmore, A.M., Falsen, E., Bahn, P., Göllner, C., Pfeffer, M., Huber, B., Busse, H.J. & Nöckler, K. (2008a)**. *Brucella* microti sp nov., isolated from the common vole *Microtus arvalis*. Int International Journal of Systematic and Evolutionary Microbiology, 58: 375–382.

97. **Scholz, H.C., Al Dahouk, S., Tomaso, H., Neubauer, H., Witte, A., Schloter, M., Kämpfer, P., Falsen, E., Pfeffer, M. & Engel, M. (2008b)**. Genetic diversity and phylogenetic relationships of bacteria belonging to the *Ochrobactrum-Brucella* group by recA and 16 S rRNA gene-based comparative sequence analysis. Systematic and Applied Microbiology, 31: 1-16.
98. **Seleem, M.N., Boyle, S.M. & Sriranganathan, N. (2008)**. *Brucella*: a pathogen without classic virulence genes. Veterinary Microbiology, 129(1-2):1-14. doi: 10.1016/j.vetmic.2007.11.023.
99. **Starr, T., Ng, T.W., Wehrly, T.D., Knodler, L.A. & Celli, J. (2008)**. *Brucella* Intracellular Replication Requires Trafficking Through the Late Endosomal/Lysosomal Compartment. Traffic, 9: 678–694.
100. **Tiller, R.V., Gee, J.E., Frace, M.A., Taylor, T.K., Setubal, J.C., Hoffmaster, A.R. & De BK. (2010)**. Characterization of novel *Brucella* strains originating from wild native rodent species in North Queensland, Australia. Applied and Environmental Microbiology, 76: 5837-5845.
101. **Tan, K.K., Tan, Y.C., Chang, L.Y., Lee, K.W., Nore, S.S., Yee, W.Y., Mat Isa, M.N., Jafar, F.L., Hoh, C.C. & AbuBakar, S. (2015)**. Full genome SNP-based phylogenetic analysis reveals the origin and global spread of *Brucella melitensis*. BMC Genomics, 18:16:93.
102. **Tsolis, R.M., Seshadri, R., Santos, R.L., Sangari, F.J., Lobo, J.M., de Jong, M.F., Ren, Q., Myers, G., Brinkac, L.M., Nelson, W.C., Deboy, R.T., Angiuoli, S., Khouri, H., Dimitrov, G., Robinson, J.R., Mulligan, S., Walker, R.L., Elzer, P.E., Hassan, K.A. & Paulsen, I.T. (2009)**. Genome degradation in *Brucella ovis* corresponds with narrowing of its host range and tissue tropism. PLoS One, 4(5):e5519. doi: 10.1371/journal.pone.0005519.
103. **Ugalde, J.E., Czibener, C., Feldman, M.F. & Ugalde, R.A. (2000)**. Identification and characterization of the *Brucella abortus* phosphoglucomutase gene: role of lipopolysaccharide in virulence and intracellular multiplication. Infection and Immunity, 68(10):5716-23.
104. **Valdezate, S., Navarro, A., Villalón, P., Carrasco, G. & Saéz-Nieto, J.A. (2010)**. Epidemiological and phylogenetic analysis of Spanish human *Brucella melitensis* strains by multiple-*locus* variable-number tandem-repeat typing, hypervariable octameric oligonucleotide fingerprinting, and *rpoB* typing. Journal of Clinical Microbiology 8: 2734–40.
105. **Velasco, J., Romero, C., López-Goñi, I., Leiva, J., Díaz, R. & Moriyón, I. (1998)**. Evaluation of the relatedness of *Brucella* spp. and *Ochrobactrum anthropi* and description of *Ochrobactrum intermedium* sp. nov., a new species with a closer relationship to *Brucella*. International journal of systematic bacteriology, 48 Pt 3:759-68.
106. **Verger, M., Grimot, D., Grimont, P. A. D. & Crayon, M. (1985)**. *Brucella* a nonspecific genus as shown by deoxyribonucleic acid hybridization. Int J Syst Bacteriol, 35: 292-295.
107. **Viadas, C., Rodríguez, M.C., Sangari, F.J., Gorvel, J.P., García-Lobo, J.M. & López-Goñi, I. (2010)**. Transcriptome analysis of the *Brucella abortus* BvrR/BvrS two-component regulatory system. PLoS One, 5(4):e10216. doi: 10.1371/journal.pone.0010216.
108. **Vizcaíno, N. & Cloeckaert, A. (2012)**. Biology and Genetics of the *Brucella* Outer Membrane, p. 133-161. In I. Lopez-Goni and D. O'Callaghan (ed.), *Brucella*: Molecular Microbiology and Genomics. Caister Academic Press, Norfolk (GBR).
109. **Watarai, M., Makino, S., Michikawa, M., Yanagisawa, K., Murakami, S. & Shirahata, T. (2002)**. Macrophage plasma membrane cholesterol contributes to *Brucella abortus* infection of mice. Infection and Immunity, 70: 4818-4825.
110. **Wattam, A.R., Foster, J.T., Mane, S.P., Beckstrom-Sternberg, S.M., Beckstrom-Sternberg, J.M., Dickerman, A.W., Keim, P., Pearson, T., Shukla, M., Ward, D.V., Williams, K.P., Sobral, B.W., Tsolis, R.M., Whatmore, A.M. & O'Callaghan, D. (2014)**. Comparative phylogenomics and evolution of the *Brucella*e: A path to virulence. Journal of Bacteriology, 196(5): 920–930. doi: 10.1128/JB.01091-13.
111. **Wattam, A.R., Inzana, T.J., Williams, K.P., Mane, S.P., Shukla, M., Almeida, N.F., Dickerman, A.W., Mason, S., Moriyón, I., O'Callaghan, D., Whatmore, A.M., Sobral, B.W., Tiller, R.V., Hoffmaster, A.R., Frace, M.A., De Castro, C., Molinaro, A., Boyle, S.M., De BK & Setubal, J.C. (2012)**. Comparative Genomics of Early-Diverging *Brucella* Strains Reveals a Novel Lipopolysaccharide biosynthesis pathway. mBio, 3(5):e00246-12. doi: 10.1128/mBio.00246-12.

112. Wattam, A.R., Williams, K.P., Snyder, E.E., N.F. Jr Almeida, M. Shukla, A.W. Dickerman, O.R. Crasta, R. Kenyon, J. Lu, J.M. Shallom, H. Yoo, T.A. Ficht, R.M. Tsolis, C. Munk, R. Tapia, C.S. Han, J.C. Detter, D. Bruce, T.S. Brettin, B.W. Sobral, S.M. Boyle and J.C. Setubal. (2009). Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. Journal of Bacteriology, 191: 3569–3579.

113. Weeks J. N., Galindo C. L., Drake K. L., Adams G. L., Garner H. R. & Ficht T. A. (2010). *Brucella melitensis* VjbR and C12-HSL regulons: contributions of the N-dodecanoyl homoserine lactone signaling molecule and LuxR homologue VjbR to gene expression. BMC Microbiology, 10: 167. doi: 10.1186/1471-2180-10-167.

114. Whatmore, A.M., Koylass, M., Muchowski, J., Edwards-Smallbone, J., Gopaul, K. & Perrett, L. (2016). Extended multi*locus* sequence analysis to describe the global population structure of the genus *Brucella*: phylogeography and relationship to biovars. Frontiers in Microbiology 7: 2049. doi:10.3389/fmicb.2016.02049.

115. Whatmore, A.M. (2009). Current understanding of the genetic diversity of *Brucella*, an expanding genus of zoonotic pathogens. Infection, Genetics and Evolution, 9: 1168–1184.

116. Whatmore, A., Perret, L. & MacMillan, A. (2007). Characterization of the genetic diversity of *Brucella* by multi*locus* sequencing. BMC Microbiology, 7: 34-48.

117. Williams, D., Gogarten, J.P. & Lapierre, P. (2011). Filling the gaps in the genomic landscape. Genome Biology, 11(2):103. doi: 10.1186/gb-2010-11-2-103.

118. Yu, W.L. & Nielsen, K. (2010). Review of Detection of *Brucella* spp. by Polymerase Chain Reaction. Croatian Medical Journal, 51(4):306-13.

Chapter 2

Isolation and molecular characterization of Brucella suis strains from swine and wild boars in Portugal

Subchapter 2.1.

Development and evaluation of a selective medium for Brucella suis

The results presented in this subchapter were previously published.

Ferreira, A.C., Almendra, C., Cardoso, R., Silva Pereira, M., Beja-Pereira, A., Luikart, G. & Corrêa de Sá, M.I. (2012). Development and evaluation of a selective medium for *Brucella suis*. *Research in Veterinary Science*, 93: 565–567.

Authors contributions:

ACF and **MICS** design the study.

ACF, CA and **RC** collected the samples.

ACF and **RC** performed the assays and interpreted the data.

ACF and **MSP** performed the statistical analysis and interpreted the data.

ACF draft the manuscript**.**

MICS, ABP and **GL** revised the manuscript.

1. Introduction

In Europe, *Brucella suis* biovar 2 is the most commonly isolated biovar in domestic pigs. It is also frequently isolated from wild boar (*Sus scrofa*) and European hares (*Lepus europaeus*), which are considered its natural reservoirs (Al Dahouk *et al.*, 2005; Cvetnic *et al.*, 2003; Godfroid and Käsbohrer, 2002; Garin-Bastuji *et al.*, 2000; Godfroid *et al.*, 1994). Wild boar seems to be the main source of infection for domestic pigs in the Iberian Peninsula (Muñoz *et al.*, 2010; EFSA, 2009). Due to the lack of specificity of serological tests, the unequivocal diagnosis depends on the isolation of *Brucella* from abortion material, udder secretions or from tissues removed at post-mortem. Likewise, bacteriological isolation although slow, expensive and cumbersome, should be performed whenever possible to confirm the disease and to determine the *Brucella* species/biovars involved (OIE, 2016). Primary isolation of the slow growing *Brucella* species requires selective media to inhibit overgrowing contaminants present in field samples. The most widely selective medium used in primary isolation is Farrell's medium (FAR; Farrell, 1974) and modified Thayer-Martin's medium (MTM; Marín *et al.*, 1996a; Brown *et al.*, 1971). FAR is an excellent medium but some antibiotics present in its composition, namely nalidixic acid and bacitracin, have inhibitory effects on some *B. suis* strains (OIE, 2016; De Miguel *et al.*, 2011; Jensen and Halling, 2010; Marín *et al.*, 1996a, 1996b). On the other hand, the MTM is suitable for isolating most brucellae and is the most effective selective medium for the isolation of *B. ovis*, *B. abortus* and *B. melitensis* but lacks inhibitory effects on many overgrowing contaminants, and hinders the assessment of colonial morphology by direct observation of culture plates (De Miguel *et al.*, 2011; Marín *et al.*, 1996a, 1996b). To overcome these problems some laboratories use both media in parallel, what makes the isolation procedure more expensive and troublesome. Therefore, in a routine laboratory that has to analyze a great amount of samples, it would be preferable to have a single selective medium for all the *Brucella* strains, without the drawbacks of the currently used media.

The objective of this work was to develop a new selective medium (named herein LNIV-M) for the isolation of *B. suis* and to assess its diagnostic performance in the isolation of *B. suis* biovar 2. It was evaluated the growth of *B. suis* reference and field strains from domestic pigs in different basal media and the susceptibility to different antibiotics contained in the currently used FAR and MTM media and the results obtained confirmed the adequate diagnostic performance of LNIV-M in the isolation of *B. suis*.

2. Material and Methods

2.1. *Brucella* strains

To select basal medium for LNIV-M, we used the reference strains, *B. suis* 1330 (biovar 1, ATCC 23444), *B. suis* Thomsen (biovar 2, ATCC 23445) and *B. suis* 686 (biovar 3, ATCC 23446) and 22 *B. suis* biovar 2 field strains from domestic pigs, identified and typed according to Alton *et al.* (1988).

2.2. Culture media

Blood Agar Base (BAB), GC medium (GC), Trypticase Soy Agar (TSA) (all from Becton–Dickinson, USA), *Brucella* Medium Base (BMB; Oxoid, England) and Plommet Medium (PM; Plommet, 1991) were evaluated as basal media by colony counting, alone or supplemented with 5% sterile horse serum (-S; Euro-Lone, Canada) or 1% haemoglobin (-H; Becton–Dickinson, USA). BMB, GC and TSA were also supplemented with 0.1% yeast extract (-YE; Becton–Dickinson, USA).

2.3. Culture conditions

Each strain was grown in TSA slants at 37 ^{o}C with 10% CO_2 for 48 h, harvested in sterile phosphate-buffered saline (PBS) pH 7.2 and spectrophotometrically (S2100 Diode Array Spectrophotometer, WPA) adjusted in the same diluent to about 10^9 CFU/ml (OD_{600} = 0.170). Then, six 10-fold dilutions of each suspension were performed in PBS and 0.1 ml of each dilution were plated by triplicate in each medium. After incubation at 37 ^{o}C for 5 days in 10% CO_2 atmosphere, the mean (n = 3) number of CFU/plate was determined in each medium, and statistically compared by ANOVA followed by LSD multiple comparison test, using the PROC GLM procedure of SAS statistical package (SAS/STAT version 9.2; SAS, 2008; Zar, 1984).

2.4. Evaluation of the relative diagnostic performance of LNIV-M in the primary isolation of *B. suis* with respect to FAR and MTM

We cultured 1649 samples (spleen; liver; lung; peripheral lymph node; reproductive organs) from 918 hunter-harvested wild boars, sampled during the hunting season (November–March) across Portugal. All tissue samples were processed as described by Alton *et al.* (1988) and 0.2 ml/plate of each tissue homogenate were cultured in duplicate plates on the three media. Samples were considered positive when at least one CFU was isolated after incubation at 37 ^{o}C (10% CO_2) for up to 10 days. *Brucella* species and biovars were identified according to Alton *et al.* (1988).

3. Results and Discussion

Our results proved that *B. suis* growth was significantly ($p < 0.0001$) inhibited in GC-S and GC-YE with respect to GC-H, BAB-S or TSA-S as described by others (De Miguel *et al.*, 2011). The later three basal media performed similarly and yielded the highest number of CFU/plate (results not shown). Considering the cost of BAB-S and that GC-H is a no-translucent medium and cumbersome to prepared, TSA-S was finally chosen as LNIV-M basal medium. In a second step, the antimicrobial supplement was formulated according to the results of Minimal Inhibitory Concentration (MIC; Sahm and Washington, 1991) and Concentration Enabling Growth (CEG; Marín *et al.*, 1996a) obtained with the 22 *B. suis* biovar 2 field strains against bacitracin, vancomycin, polymyxin B, nalidixic acid, cycloheximide, nystatin, colistin, nitrofurantoin and amphotericin B (all from Sigma–Aldrich, Germany), in the range of 512–1 mg/L, using Mueller–Hinton broth and agar (Becton–Dickinson, USA). Results are presented in **Table 2.1.1**. The variations in susceptibility to the antibiotics were notorious, except for amphotericin B and cycloheximide antifungal agents which, as expected, did not affect the growth of *B. suis* bv. 2. In particular, most of the strains (13/22; 59.1%) were inhibited by low concentrations of polymyxin B (MIC ≤4 mg/L and CEG ≤1 mg/L), but all are resistant to higher concentrations of colistin (MIC ≥16 mg/L and CEG ≥4 mg/L). These results are in agreement with those obtained by De Miguel and colleagues (2011).

Considering the MIC and CEG results (**Table 2.1.1**), LNIV-M was formulated with TSA-S supplemented with 20 mg/L vancomycin (bactericidal for most gram positive bacteria), 4 mg/L colistin and 16 mg/L of nitrofurantoin (active against most gram negative bacteria), and 100 mg/L cycloheximide and 16 mg/L (95 000 IU/L) nystatin (active against yeasts and moulds). The efficacy of LNIV-M for culturing *B. suis* was compared with that of FAR and MTM media (both prepared as described elsewhere; Marín *et al.*, 1996a; Farrell, 1974), using the *B. suis* reference and field strains described above. The number of CFU/plate grew in LNIV-M, FAR and MTM selective medium and TSA-S (control) was determined and statistically compared as detail above. The overall results are shown in **Table 2.1.2**. No significant differences were found for LNIV-M and MTM with respect to the TSA-S medium used as control, but FAR significantly ($p < 0.05$) reduced the number of CFU. These data confirm our preliminary studies and the results obtained for other *Brucella* species (De Miguel *et al.*, 2011; Marín *et al.*, 1996a, 1996b). Moreover, colonial size on LNIV-M and MTM were smaller than on TSA-S but larger than on FAR. Since MTM contains hemoglobin, the translucent LNIV-M has the advantage to allow a better direct identification of *Brucella*.

Table 2.1.1. Minimal Inhibitory Concentration (MIC) and Concentration Enabling Growth (CEG) for 22 *B. suis* field strains

Antibiotics	No. of strains	MIC (mg/L)	CEG (mg/L)
Bacitracin	18	512	256
	4	256	128
Vancomycin	2	>512	512
	10	512	256
	8	512	128
	2	256	64
Colistin methanosulfonate	7	64	16
	12	32	8
	3	16	4
Polymixin B sulphate	9	64	16
	11	4	1
	2	1	<1
Nalidixic Acid	19	64	32
	3	32	16
Nitrofurantoin	8	512	256
	6	512	128
	6	512	64
	2	512	32
Amphotericin B	22	>512	512
Cycloheximide	21	512	256
	1	256	128
Nystatin	7	512	256
	9	256	64
	5	128	32
	1	64	16

Table 2.1.2. Susceptibility of *B. suis* reference and field strains to the new LNIV-M selective culture medium, in comparison to modified Thayer-Martin (MTM) and Farrell (FAR) media.

Culture media	CFU/plate (mean ± SD) of *B. suis**			
	Reference strains			Field strains
	biovar 1 ATCC 23444	biovar 2 ATCC 23445	biovar 3 ATCC 23446	biovar 2 (n=22)
TSA-S (control)	84.0 ± 7.81 [B]	75.0 ± 15.10 [B]	78.7 ± 5.51 [B]	72.1 ± 21.57 [B]
LNIV-M	77.3 ± 8.08 [B]	83.3 ± 6.66 [B]	81.3 ± 1.53 [B]	67.0 ± 20.77 [B]
MTM	68.3 ± 15.04 [B]	73.3 ± 7.57 [B]	46.0 ± 14.00 [B]	64.4 ± 23.12 [B]
FAR	22.3 ± 2.08 [A]	25.7 ± 3.06 [A]	21.0 ± 1.00 [A]	48.6 ± 18.74 [A]

***** Mean and SD (n=3 from independent experiments with the corresponding reference strain; or n=22 field strains) of the number of CFU/plate were determined, after triplicate plating of 0.1 ml of a suspension containing around 10^3 CFU/ml in each culture medium. TSA-S was used as control.

[A,B] Values with the same letter indicate that means were statistically equivalent (p>0.05)

Finally, to evaluate the relative diagnostic performance of LNIV-M in the primary isolation of *B. suis* with respect to FAR and MTM, we cultured 1649 samples (spleen; liver; lung; peripheral lymph node; reproductive organs) from 918 hunter-harvested wild boars, sampled during the hunting season (November–March) across Portugal. All tissue samples were processed as described by Alton *et al.* (1988) and 0.2 ml/plate of each tissue homogenate were cultured in duplicate plates on the three media. Samples were considered positive when at least one CFU was isolated after incubation at 37 ^{o}C (10% CO_2) for up to 10 days. *Brucella* species and biovars were identified according to Alton *et al.* (1988). The number of plates rejected due to overgrowing contaminants in LNIV-M was lower than in MTM but higher than in FAR (data not shown). From the 918 wild boars tested, 63 animals (6.9%) were found to be infected with *B. suis* biovar 2, showing a total of 139 (8.4%) positive samples (63 spleens; 21 livers; 21 lymph nodes; 21 lungs; 13 reproductive organs). LNIV-M detected 59 out the 63 (93.6%) positive animals and 87 out the 139 (62.6%) positive samples, while FAR and MTM detected, respectively, 58 (92.1%) and 50 (79.4%) of animals, and 81 (58.3%), and 83 (59.7%) of samples. The combined use of LNIV-M and MTM detected 95.2% (60) positive animals, whereas LNIV-M and FAR detected 100% positive animals. The results obtained in the diagnostic performance of FAR and MTM are not in agreement with those from efficacy studies. This can be explained by the high inhibitory effect of FAR for most contaminants present in field samples, allowing the growth of *Brucella* colonies and showing that this medium is more suitable for diagnostic purpose than MTM. The Cohen's Kappa coefficient (K) and McNemar's Chi-square test (X^2; Zar, 1984) were used to calculate the level of agreement between media. A p less than 0.05 ($p < 0.05$) was considered statistically significant. Analysis of samples showed a substantial agreement between media, except for lung tissue, where poor recovery of *Brucella* was achieved. The best value was observed when comparing LNIV-M and FAR ($K = 0.699$, 95% CI; $X^2 = 1.25$, $p \geq 0.05$, data not shown).

4. Conclusion

In conclusion, for an adequate bacteriological diagnosis of brucellosis and to increase its sensitivity, more than one selective culture medium should be used. Data obtained in this study indicate that LNIV-M is a useful selective medium for isolation of *B. suis* and the combined use of LNIV-M and FAR has shown to improves the performance of the bacteriological diagnosis of brucellosis in swine.

5. References

1. **Al Dahouk, S., Nöckler, K., Tomaso, H., Splettstoesser, W.D., Jungersen, G., Riber, U., Petry, T., Hoffmann, D., Scholz, H.C., Hensel, A. & Neubauer, H. (2005)**. Seroprevalence of Brucellosis, Tularaemia and Yersiniosis in Wild Boars (*Sus scrofa*) from North-Eastern Germany. Journal of Veterinary Medicine, 52: 444–455.
2. **Alton, G.C., Jones, L.M., Angus, R.D. & Verger, J.M. (1988)**. Techniques for the Brucellosis Laboratory. Institute National de la Recherche Agronomique, Paris, France.
3. **Brown, G.M., Ranger, C.R. & Kelly, D.J. (1971**). Selective media for the isolation of *Brucella* ovis. Cornell Veterinarian, 61:265–280.
4. **Cvetnic, Z., Mitak, M., Ocepek, M., Lojkic, M., Terzic, S., Jemersic, L., Humski, A., Habrun, B., Sostaric, B., Brstilo, M., Krt, B. & Garin-Bastuji, B. (2003)**. Wild boars (*Sus scrofa*) as reservoirs of *Brucella suis* biovar 2 in Croatia. Acta Veterinaria Hungaria, 51: 465–473.
5. **De Miguel, M.J., Marín, C.M., Muñoz, P.M., Dieste, L., Grilló, M.J. & Blasco, J.M. (2011**). Development of a selective medium for primary isolation of the main *Brucella* species. Journal of Clinical Microbiology, 49: 1458–1463.
6. **EFSA (2009**). Scientific Opinion of the Panel on Animal Health and Welfare (AHAW) on a request from the Commission on porcine brucellosis (*Brucella* suis). The EFSA Journal, 1144: 1–112.
7. **Farrell, I.D. (1974)**. The development of a new selective medium for the isolation of *Brucella abortus* from contaminated sources. Research in Veterinary Science, 16: 280–286.
8. **Garin-Bastuji, B., Hars, J., Calvez, D., Thiébaud, M. & Artois, M. (2000**). Brucellose du porc domestique et du sanglier sauvage due à *Brucella suis* biovar 2 en France. Epidémiologie and Santé Animale, 38: 1–5.
9. **Godfroid, J. & Käsbohrer, A. (2002**). Brucellosis in the European Union and Norway at the turn of the twenty-first century. Veterinary Microbiology, 90: 135–145.
10. **Godfroid, J., Michel, P., Uytterhaiegen, L., De Smedt, C., Rasseneur, F., Boelaert, F., Saegerman, C. & Patigny, X. (1994)**. Brucellose enzootique a *Brucella suis* biotype 2 chez le sanglier (*Sus scrofa*) en Belgique. Les Annales de Médicine Vétérinaire 138, 263–268.
11. **Jensen, A.E. & Halling, S.M. (2010**). Effect of polymyxin B and environmental conditions on isolation of *Brucella* species and the vaccine strain RB51. Comparative Immunology, Microbiology and Infectious Diseases, 33: 121–131.
12. **Marín, C.M., Alabart, J. & Blasco, J.M. (1996a**). Effect of antibiotics contained in two *Brucella* selective media on growth of *B. abortus*, *B. melitensis* and *B. ovis*. Journal of Clinical Microbiology, 34: 426–428.
13. **Marín, C., Jimenez de Bagüés, M.P., Barberán, M. & Blasco, J.M. (1996b)**. Comparison of two selective media for the isolation of B. melitensis from naturally infected sheep and goats. Veterinary Record, 138: 409–411.
14. **Muñoz, M.P., Boadella, M., Arnal, M.C., De Miguel, M., Revilla, M., Martinez, D., Vicente, J., Acevedo, P., Oleaga, A., Ruiz-Fons, F., Marin, C., Prieto, J.M., De la Fuente, J., Barral, M., Barberan, M., De Luco, D., Blasco, J.M. & Gortazar, C. (2010)**. Spatial distribution and risk factors of Brucellosis in Iberian wild ungulates. BMC Infectious Diseases, 10: 46.
15. **Plommet, M. (1991**). Minimal requirements for growth of *Brucella suis* and other *Brucella* species. Zentralblatt fur Bakteriologie, 275: 436–450.
16. **Sahm, D.F. & Washington, J.A. (1991**). Antibacterial susceptibility tests: dilution methods. In: Balows, A., Hausler, W.J., Jr., Hermann, K.L., Isenberg, H.D., Shadomy, H.J. (Eds.), Manual of Clinical Microbiology, fifth ed. American Society for Microbiology, Washington, D.C., pp. 1105–1116.
17. **SAS Institute (2008)**. SAS/STAT Version 9.2, User's Guide, second ed. SAS Institute Inc., Carry, NC, p. 208.
18. **Zar, J.H. (1984)**. Biostatistical Analysis, second ed. Prentice Hall International, London.

Subchapter 2.2.

Genetic diversity of Brucella suis biovar 2 strains circulating in Europe

The results presented in this subchapter were submitted for publication to *Veterinary Microbiology*.

Ferreira, A.C., Tenreiro, R., Dias, R. & Corrêa de Sá, M.I. Genetic diversity of *Brucella suis* biovar 2 strains circulating in Europe.

Authors contributions:

ACF and **MICS** design the study, performed all the molecular genotyping methods and draft the manuscript.

ACF, RD and **RT** analyzed the data.

RT revised the manuscript.

1. Introduction

Brucella suis consists of five biovars (named from 1 to 5) that affects a wide range of hosts including man. The infection in swine is caused by biovars 1, 2, and 3. *B. suis* has received meager attention comparatively to other *Brucella* species and, despite the greater internal diversity and wide host range, studies have been focused on biovar 1 (EFSA, 2009). The infection caused by biovar 2 differs from 1 and 3 in its host range, its limited geographical distribution and its pathology. Infections due to *B. suis* biovar 1 and 3 have been reported worldwide in several animal species and humans. Brucellosis due to biovar 2 represents an emerging disease in domestic swine throughout Europe and is associated with the increase of extensive swine farms and the high density of infected wild boars, representing an important hazard particularly for the Iberian pig population reared in outdoor breeding systems (Muñoz *et al.*, 2010; EFSA, 2009). This infection has been placed in evidence in Belgium (Grégoire *et al*, 2012), Croatia (Cvétnic *et al.*, 2005), Czech Republic (Hubalek *et al.*, 2002), Denmark (Jungersen *et al.*, 2006), France (Garin-Bastuji *et al.*, 2000), Germany (Al Dahouk *et al.*, 2005), Hungary (Hubalek, *et al.*, 2002), Italy (De Massis *et al.*, 2012), Poland (Szulowski *et al.*, 2013), Portugal (Ferreira *et al.*, 2012), Romania (EFSA, 2009), Serbia and Montenegro (Djuricic, 2010), Spain (Muñoz *et al.*, 2010) and Switzerland (Koppel *et al.*, 2007). In contrast, *B. suis* has never been found in Finland, Norway, Sweden and the United Kingdom (EFSA, 2009).

The identification of *B. suis* biovars is usually performed by standard bacteriological methods. However, these tests lack specificity and are not straightforward particularly for the identification of biovars 1, 2, and 3 (Ferrão-Beck *et al.*, 2006). Since accurate typing procedures are essential for epidemiological studies, different PCR-based methods have been proposed to improve *B. suis* molecular characterization. Suis-ladder multiplex PCR is a rapid and robust system that allows a fast and precise identification of *B. suis* at the biovar level (López-Goñi *et al.*, 2011). The PCR-restriction fragment length polymorphism (PCR-RFLP) analysis of genes *omp2a* and *omp2b* (Cloeckaert *et al.*, 1995), and gene *omp31* (Vizcaíno *et al.*, 1997) showed to be useful for differentiation between biovars and revealed additional polymorphism within biovar 2 (Muñoz *et al.*, 2010; Garcia-Yoldi *et al.*, 2007; Ferrão-Beck *et al.*, 2006). Molecular typing by Multi*locus* Variable Number Tandem Repeat Analysis (MLVA) is an accurate method to determine the relatedness amongst bacterial isolates and has been frequently used to disclose genetic relationships inside *Brucella* genus, at species, biovar and strain level (Scholz & Vergnaud, 2013). In fact, the *Brucella* MLVA-16 assay, originally developed by Le Flèche *et al.* (2006) and modified by Al Dahouk *et al.* (2007), consists of 16 genetic markers comprising eight minisatellite markers most appropriate for species-level identification, and eight microsatellite markers with higher discriminatory power. This assay is easily performed and reproducible, and

demonstrated the high heterogeneity at the genomic level between *B. suis* isolates (Duvnjak *et al.*, 2015; Kreizinger *et al.*, 2014; Li *et al.*, 2013; Garcia-Yoldi *et al.*, 2007).

In this work, different PCR-based methods, including the Suis-ladder multiplex PCR, PCR-RFLP analysis for *omp2a, omp2b* and *omp31* genes and MLVA-16 assay, were used to assess the biovar, the different haplotypes and the epidemiological relationship between strains from diverse host and geographic origins, including Portugal and other European countries, in order to disclose the genetic diversity of *B. suis* biovar 2 strains circulating in Europe.

2. Material and Methods

2.1. *B. suis* isolates and genomic DNA bank

One hundred and seventy six *B. suis* isolates were used in this study, comprising 11 isolates from biovar 1, 160 from biovar 2, one from biovar 3 and four from biovar 4. Isolates were obtained from diverse geographic origins and hosts (**Table 2.2.1**), including human (n=4), bovine (n=1), caribou (n=1), goat (n=2), hare (n=6), reindeer (n=3), sheep (n=2), swine (n=62) and wild boar (n=94). Reference strains from biovar 1 (strain 1330, ATCC 23444), biovar 2 (strain Thomsen, ATCC 23445), biovar 3 (strain 686, ATCC 23446), biovar 4 (strain 40, ATCC 23447) and biovar 5 (strain 513, NCTC 11996) were also included in the study. All *B. suis* isolates were typed according to standard bacteriological procedures (Alton *et al.*, 1988), and a bank of genomic DNA extracts for all isolates and reference strains was prepared using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany), according to the manufacturer, and conserved at -80ºC.

2.2. Suis ladder multiplex PCR and PCR-RFLP analysis

All *B. suis* isolates were subjected to the Suis-ladder multiplex PCR (López-Goñi *et al.*, 2011) and PCR-RFLP analysis for *omp2a* and *omp2b* (Cloeckaert *et al.*, 1995), and *omp31* (Vizcaíno *et al.*, 1997) genes to accurately assess the biovar and the different haplotypes, as previously described. All DNA primers were purchased from Invitrogen and restriction enzymes were from Biolabs (*Eco*RI and *Nco*I) or Promega (AvaII, *Hae*III, *Sty*I and *Kpn*I).

2.3. MLVA-16 assay and data analysis

Single-*locus* amplifications of the 16 genetic markers were performed as described elsewhere (Ferreira *et al.*, 2012). The 16 *loci* have been classified in three panels, named panel 1, composed of 8

minisatellite (bruce06, bruce08, bruce11, bruce12, bruce42, bruce43, bruce45 and bruce55), panel 2A (bruce18, bruce19 and bruce21) and panel 2B (bruce04, bruce07, bruce09, bruce16 and bruce30) composed of three and five microsatellite markers, respectively. Briefly, PCR reactions were performed in a total volume of 15 µl containing 3 ng of DNA, 1X PCR Reaction Buffer, 1 U of Taq DNA polymerase (Roche Diagnostics, Germany), 200 µM of each dNTPs and 0.3 µM of each flanking primers. Amplifications were performed in a MyCycler thermal Cycler (Biorad, France). An initial denaturation step at 96 ^{o}C for 5 min was followed by 30 cycles of denaturation at 96 ^{o}C for 30 s, primer annealing at 60 ^{o}C for 30 s and elongation at 70 ^{o}C for 1 min. The final extension step was performed at 70 ^{o}C for 5 min. Five microliters of amplification products were loaded on a 3% standard agarose gel to analyze panel 2A and 2B *loci* (tandem repeats with a unit length shorter than 8 bp) and on a 2% standard agarose gel for panel 1 *loci* (tandem repeats with a unit length larger than 10 bp) and run under a voltage of 8 V/cm for 120 min. Depending on the tandem repeat unit length, a 20 bp (20 bp PCR Molecular Ruler, Biorad, France) or a 100 bp ladder was used as molecular size marker (100 bp DNA ladder, Invitrogen). The total number of repeats at each *locus* was determined by the correlation with the amplicon size according to the 2013 *Brucella* allele assignment table (Le Flèche *et al.* 2006 version 3.6 available at http://mlva.u-psud.fr). Genomic DNA from *B. melitensis* biovar 1 strain 16M (ATCC 23456) and *B. suis* biovar 2 strain Thomsen were used as controls for alleles assignment. Cluster analysis of MLVA-16 data was based on the categorical coefficient (with equal weight for all markers) and unweighted pair group method with arithmetic averages (UPGMA) using BioNumerics version 6.5 (Applied Maths, Belgium). The MLVA-16 genotypes of the reference strains were also included in the analysis and a standard minimum spanning tree (MST) was applied to the set of 181 strains. MLVA data from this work was compared to results obtained by other authors with *B. suis* strains from different origins that were deposited in the *Brucella* MLVA database (available at http://mlva.u-psud.fr). The genetic diversity of the *loci* was calculated using the Hunter–Gaston diversity index (HGDI) (Hunten & Gaston, 1988), via the online tool V-DICE available at the HPA website (http://www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl).

A similar MST analysis of MLVA-11 genotypes (including only mini-/microsatellite markers from panels 1 and 2A) was also performed for a total of 526 strains, including 345 additional *B. suis* strains from different biovars, using the data available in the *Brucella* database hosted at http://microbesgenotyping.i2bc.paris-saclay.fr.

Table 2.2.1. *Brucella suis* isolates used in the study.

Species /biovar	Country	Number of isolates	Hosts [a]
B. suis biovar 1	Croatia (HR)	2	G; SW
	France (FR)	5	H; M; SW
	Mexico (MX)	1	M
	Netherlands (NL)	1	M
	French Polynesia (PF)	1	SW
	United States (US)	1	SW
	Total	11	
B. suis biovar 2	Belgium (BE)	20	SW
	Croatia (HR)	1	SW
	France (FR)	10	G; H; SW; WB
	Denmark (DK)	1	n.a.
	Germany (DE)	5	H; WB
	Italy (IT)	7	WB
	Portugal (PT)	92	B; SW; S; WB
	Spain (SP)	22	SW; WB
	Switzerland (CH)	1	WB
	Unknown	1	SW
	Total	160	
B. suis biovar 3	India (IN)	1	M
B. suis biovar 4	Alaska	1	C
	Poland (PL)	3	R
	Total	4	

[a] B, bovine; C, caribou; G, goat; H, hare; M, man; R, reindeer; S, sheep; SW, Swine; WB, wild boar; n.a., information not available.

3. Results and Discussion

The genetic structure of 176 *B. suis* isolates and five reference strains representative of each biovar was characterized using the Suis-ladder multiplex PCR, the PCR-RFLP analysis for *omp2a*, *omp2b* and *omp31* genes and the MLVA-16 assay. Full information on *B. suis* isolates, including geographic origin and molecular typing data obtained with Suis-ladder multiplex PCR and RFLP-PCR is detailed in **Supplementary Table S2.2.1**. A perfect correlation was found between classical typing and Suis-ladder multiplex PCR, since all isolates allocated to a particular biovar displayed the expected specific band profile according to López-Goñi *et al.* (2011). Likewise, PCR-RFLP analysis of *omp2a, omp2b* and *omp31* genes also detected different haplotypes according to the specific biovar (Vizcaíno *et al.*, 1997; Cloeckaert *et al.*, 1995). However, additional polymorphism within biovar 2 isolates was identified, describing five different restriction patterns named 2a, 2b, 2c, 2d and 2e. Haplotype 2a was characteristic of the *B. suis* biovar 2 Thomsen reference strain and was found in hare and swine isolates from France and Germany; haplotype 2b was only observed in four isolates from France (three from

hare and one from swine); and haplotype 2c was observed in 36 swine and wild boar isolates from Belgium, Croatia, France, Italy, Spain and Switzerland, as well as in one goat isolate from France. Haplotypes 2d and 2e were restricted to Iberian Peninsula. In Portugal, 79.3% (73/92) of the isolates (including swine, wild boar, bovine and sheep as hosts) corresponded to haplotype 2d and 20.7% (19/92) were haplotype 2e (only found in isolates from wild boar). Considering the Spanish isolates, haplotype 2d (54.5%; 12/22) and haplotype 2e (31.8%; 7/22) were both present in isolates from swine and wild boar. A DNA-based study in isolates from wild boar performed by Muñoz and colleagues (2010) have also shown that haplotype 2d is the most frequent in Spain in both swine and wild boar.

Identification of species and biovars of *Brucella* field strains isolated from outbreaks, paired with a genotyping approach such as MLVA, is essential to fully understand the epidemiology of the disease and to trace sources of infection, as well as to detect the introduction of any new strains. MLVA-16 has been applied to evaluate the epidemiological relationships among *Brucella* spp. isolates from different geographical origins, showing to be extremely discriminant and highly efficient to confirm epidemiological linkage in outbreak investigations. Nevertheless, most of the studies were applied to *B. abortus* or *B. melitensis* (Sun *et al.*, 2016; Allen *et al.*, 2015; De Massis *et al.*, 2015; Shevtsov *et al.*, 2015; Tay *et al.*, 2015; Garafolo *et al.*, 2013; Jiang *et al.*, 2013; Ferreira *et al.*, 2012; Aftab *et al.*, 2011; Kiliç *et al.*, 2011; Valdezate *et al.*, 2010; Al Dahouk *et al.*, 2007; Le Flèche *et al.*, 2006) and fewer studies targeted the epidemiological context of *B. suis* (Duvnjak *et al.*, 2015; Tay *et al.*, 2015; Kreizinger *et al.*, 2014; Jiang *et al.*, 2013; Li *et al.*, 2013; Garcia-Yoldi *et al.*, 2007). In this report, MLVA-16 was performed to study the genetic diversity of 176 *B. suis* isolates, and the results of this investigation represent the first data concerning the *B. suis* biovar 2 genotypes (GTs) circulating in Portugal (**Supplementary Table S2.2.2**).

To assess the diversity of *B. suis* populations and to compare the discriminatory power of MLVA typing approaches, HGDI values were calculated for all individual markers, as well as for the combined MLVA-16, MLVA-11 (panels 1 and 2A markers) and MLVA-8 (panel 1 markers) subsets (**Table 2.2.2**). All 16 *loci* have been identified for all strains except bruce09 in two swine isolates from France. Two markers (bruce45 and bruce21) were monomorphic in all biovars, two others (bruce06 and bruce43) only showed a different repeat copy number in biovar 5 and a moderate discrimination was found in bruce18 and bruce55 ($0.60 \leq HGDI \leq 0.70$). In contrast, bruce09, bruce04, bruce30 and bruce07 from panel 2B were highly discriminatory ($HGDI > 0.80$). Although bruce16 is usually included in the more variable panel 2B (Le Flèche *et al.*, 2006), it showed a low diversity index for *B. suis* (0.20). This was due to the fact that this *locus* is monomorphic for biovar 2 isolates, results that were also obtained in previous studies encompassing isolates from Spain (Garcia-Yoldi *et al.*, 2007), Hungary (Kreizinger *et*

al., 2014) and Croatia (Duvnjak *et al.*, 2015). It is also noteworthy that bruce12 and bruce42 from panel 1 clearly differentiated isolates with PCR-RFLP haplotypes 2d and 2e from the others (**Table 2.2.2**).

The dendrogram of the genetic relatedness of all *B. suis* strains is depicted in **Figure 2.2.1** (A and B). MLVA-16 distinguished a total of 126 GTs with 101 singleton GTs. The VNTR profiles are shown in **Supplementary Table S2.2.2** and were uploaded in the *Brucella* MLVA database at http://mlva.u-psud.fr. *B. suis* represented an heterogeneous group (global similarity of 21.2%), showing an extreme diversity among strains (HGDI = 0.99), mainly due to diverging panel 2B VNTR markers. However, MLVA-16 was highly consistent among isolates with known epidemiological links and sharing the same MLVA-16 GTs (**Supplementary Table S2.2.2**). Considering a similarity cutoff value of 40%, *B. suis* isolates were grouped into three major clusters. Cluster I grouped all isolates from biovar 1, 3 and 4 and the respective reference strains, with biovar 3 more closely related to biovar 1 than to biovar 4. Cluster II included all biovar 2 isolates and was further separated into two sub-clusters (IIA and IIB), using a cutoff value of 50% similarity. Sub-cluster IIA included the biovar 2 reference strain (isolated in Denmark), all isolates from Central-European countries (Belgium, Croatia, France, Denmark, Germany, Italy and Switzerland) and 3 isolates from Spain, sharing PCR-RFLP haplotypes 2a, 2b and 2c. All Portuguese isolates and the remaining Spanish isolates, encompassing haplotypes 2d and 2e, were grouped in sub-cluster IIB. Lastly, *B. suis* biovar 5 reference strain formed a separate simplicifolious line (cluster III). A MST was also constructed to display the relationships of the various MLVA-16 types (**Figure 2.2.1C**). A considerable geographical structure was found among isolates and two clonal lineages can be defined for biovar 2 isolates: the Central-European clonal lineage (sub-cluster IIA; haplotypes 2a, 2b and 2c) and the Iberian clonal lineage (sub-cluster IIB; haplotypes 2d and 2e), with Iberian clonal lineage isolates more distant from biovar 1, 3 and 4 (cluster I) than Central-European isolates. It is also noteworthy that biovar 5 (cluster III), although distinct from other *B. suis* biovars, may be more closely related to biovar 2 isolates from Iberian clonal lineage. The results obtained in this report were consistent with those previously identified by other authors (Muñoz *et al.*, 2010; Garcia-Yoldi *et al.*, 2007; Ferrão-Beck *et al.*, 2006).

Table 2.2.2. Repeat copy numbers at each *locus* in the MLVA-16 assay and Hunter-Gaston Diversity Index (HGDI) for each *locus* and MLVA-16, MLVA-11 and MLVA-8 subsets.

VNTR markers[a]	No. of tandem repeat copies at each *locus* in the following PCR-RFLP haplotypes[b]									HGDI	CI 95% [d]
	1a	2a	2b	2c	2d	2e	3a	4a	5a[C]		
Panel 1											
Bruce06	2	2	2	2	2	2	2	2	1	0.01	0.00 – 0.03
Bruce08	3;4	4;5	3-5	3-5	5	5	3	3	2	0.49	0.42 – 0.56
Bruce11	6;8	8;9	8	8	8	8	4	9	9	0.21	0.14 – 0.29
Bruce12	10	14	14	14	9	9	11	10	16	0.54	0.48 – 0.60
Bruce42	4	6	6	6	5	5	3	3	1	0.54	0.48 – 0.60
Bruce43	1	1	1	1	1	1	1	1	2	0.01	0.00 – 0.03
Bruce45	5	5	5	5	5	5	5	5	5	0.00	n.a.
Bruce55	2	2;7	2	6;7	5	4	2	2	5	0.68	0.63 – 0.73
Panel 2A											
Bruce18	4-6	4;6	4;6;7	4-6	5-7	5-7	4	4;5	7	0.62	0.58 – 0.67
Bruce19	38;43	43	38;43	38;43	38	38	38	36;40	44	0.44	0.38 – 0.50
Bruce21	9	9	9	9	9	9	9	9	9	0.00	n.a.
Panel 2B											
Bruce04	5-7	6;8;9;12;14;20	4;11;13;14	4;6-16;20;22;23	3;5;8-15;17;20;21	7;9;11-13;20	7;8	4;5;8;12	9	0.92	0.90 – 0.93
Bruce07	5-7	9;10;12;14	5;7;12	5-10	4-10;15	5;7;9-15	5;6	5;6;8	5	0.83	0.80 – 0.86
Bruce09	5-7;10	10;12;16;18-20	3;5;9;21	3;7;9-12;15-19;22	7-21	7-11;13;15;17	10;12	9;10;16	3	0.93	0.93 – 0.94
Bruce16	4-6	2	2	2	2	2	4	5;6;9	9	0.20	0.13 – 0.28
Bruce30	3;4	4;6;11	4;5	5;7-12	5-10	4;5;7;8	5;7	3;5	5	0.85	0.83 – 0.87
MLVA-8 (16 genotypes)										0.71	0.65 – 0.77
MLVA-11 (26 genotypes)										0.86	0.83 – 0.89
MLVA-16 (126 genotypes)										0.99	0.98 – 1.00

[a] MLVA-16 markers are defined as described in Le Flèche *et al.*, 2006 and Al Dahouk *et al.*, 2007.

[b] PCR-RFLP haplotypes are defined as described in Material and Methods section and shown in Additional file 1.

[c] Only the genotype of the reference strain *B. suis* biovar 5 is considered.

[d] n.a.: not applicable.

MLVA-11 subset is usually used to better describe epidemiological linkages of genotypes with the geographic origin (De Massis *et al.*, 2015; Shevtsov *et al.*, 2015; Kreizinger *et al.*, 2014; Li *et al.*, 2013; Ferreira *et al.*, 2012; Valdezate *et al.*, 2010; Le Flèche *et al.*, 2006). In this work, MLVA-11 discriminated 26 genetic variants that were compared with typing data published by other authors (Duvnjak *et al.*, 2015; Kreizinger *et al.*, 2014; Jiang *et al.*, 2013; Garcia-Yoldi *et al.*, 2007; Le Flèche *et al.*, 2006). Two MLVA-11 GTs were recorded for the first time for biovar 1 isolates, eight for biovar 2 and one for biovar 4 (**Supplementary Table S2.2.2**). Even though the variability observed in the number of GTs, all variants involved only one or two *loci*.

Considering the biovar 2 population, 12 MLVA-11 GTs were assigned to Central-European cluster IIA isolates (n=48), from which seven were novel (GT236, GT237, GT239, GT244, GT246, GT253 and GT254) and five (GT41, GT44, GT45, GT48 and GT50) were formerly described (Kreizinger *et al.*, 2014; Le Flèche *et al.*, 2006). Although GT244 was formally assigned in this work, it was previously observed in one wild boar isolate from Switzerland by Kreizinger *et al.* (2014). In this cluster, only two isolates, one from hare and one from unknown host, shared the same GT50 as biovar 2 reference strain, and the most predominant genotype was GT44 (41.7%; 20/48), found in isolates from Belgium, France, Italy and Switzerland. Less variability was found in *B. suis* Iberian isolates (cluster IIB; n=111), which were allocated only to five different specific GTs: one GT recorded for the first time (GT233) and four (GT58, GT59, GT60 and GT61) previously described by other authors (Kreizinger *et al.*, 2014; Le Flèche *et al.*, 2006). The MLVA-11 GT58 was assigned to 47.7% (53/111) of the Iberian isolates, suggesting the existence of an ongoing colonization of Iberian Peninsula with this specific MLVA-11 lineage. The lower diversity of the Iberian lineage is also evident from the less variable MLVA panel 1 markers (MLVA-8), since only two specific GTs (GT18 and GT19) were found in Iberian Peninsula, while eight GTs were assigned to isolates from Central-European countries.

In order to get additional insights into the evolutionary associations between *B. suis* lineages and their host species of origin, a MST analysis of 526 strains was performed using the MLVA-11 typing data (**Figure 2.2.2**), including those from this study and others previously published (Duvnjak *et al.*, 2015; Kreizinger *et al.*, 2014; Jiang *et al.*, 2013; Garcia-Yoldi *et al.*, 2007; Le Flèche *et al.*, 2006) and available in the *Brucella* database hosted at http://microbesgenotyping.i2bc.paris-saclay.fr. The analysis included 85 strains from biovar 1, 422 from biovar 2, four from biovar 3, 11 from biovar 4, three from biovar 5 and one rough strain.

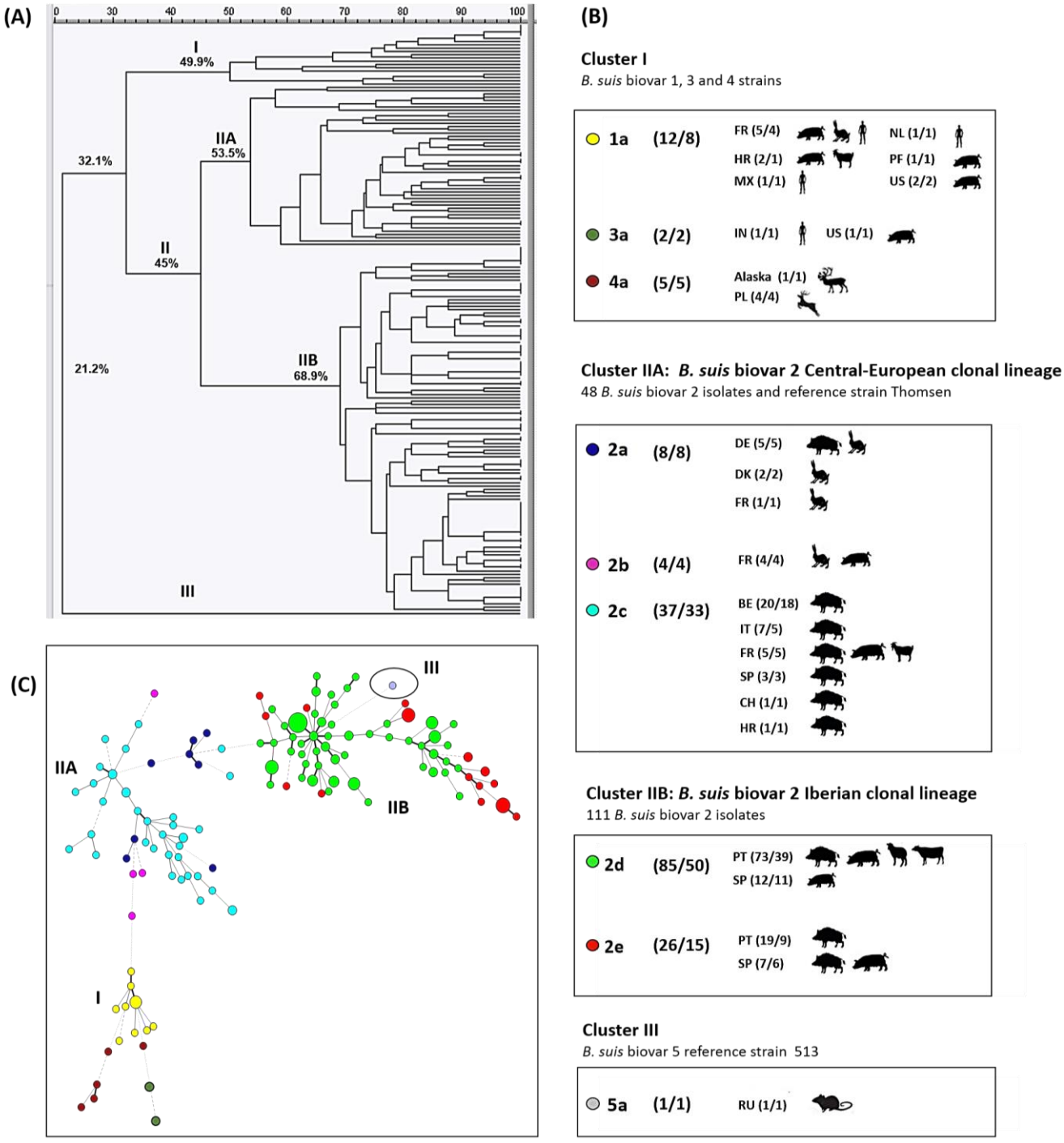


Figure 2.2.1. MLVA-16 clustering analysis. **(A)** MLVA-16 dendrogram of 176 *B. suis* isolates compared to the reference strains. Cluster analysis was performed based on the categorical coefficient and UPGMA. Cut-off similarity values were defined at 40% similarity for major clusters (I to III) and 50% similarity for sub-clusters (IIA and IIB). **(B)** Frequency distribution of all isolates in clusters I, IIA, IIB and III. The PCR-RFLP haplotypes (1a, 2a-2e, 3a, 4a and 5a), the number of isolates by number of MLVA-16 genotypes, the country of origin and host species are shown. Colored circles code to PCR-RFLP haplotypes. Country of origin: Alaska, Belgium (BE), Croatia (HR), Denmark (DK), Germany (DE), India (IN), Italy (IT), Mexico (MX), Netherlands (NL), Poland (PL), Portugal (PT), Polynesia (PF), Spain (SP), Switzerland (CH), United States (USA) and former USSR (RU; reference strain *B. suis* 513). **(C)** MLVA-16 minimum spanning tree describing the relationships of *B. suis* isolates. Circles represent MLVA-16 genotypes, colored according to PCR-RFLP haplotypes, and the size of the circle indicates the number of strains with that genotype. Dendrogram clusters (I, IIA, IIB and III) are also pointed out.

The MLVA-11 MST topology (**Figure 2.2.2A**) is similar to the one found with MLVA-16 analysis (**Figure 2.2.1C**), clearly depicting the Iberian and Central-European clonal lineages in biovar 2 and the clustering of strains from biovars 1, 3 and 4. Regarding biovar 5, a more close relation to biovar 2 Central-European clonal lineage was found with the less variable markers of MLVA-11. Beyond biovar associations, also host species associations can be observed in the MST analysis (**Figure 2.2.2B**). In biovar 2 Iberian clonal lineage, the prevailing host species are swine (59.6%; 99/166) and wild boar (38.6%; 64/166); in contrast, in biovar 2 Central-European clonal lineage, hare seems to be a relevant host (18.4% of the strains; 47/256), along with the most common hosts swine (34.4%;88/256) and wild boar (43.4%; 111/256). The majority of biovar 1 strains (50.6%; 43/85) were isolated from swine, although isolates from hare (8.2%; 7/85) and wild boar (3.5%; 3/85) were observed. The *B. suis* collection also includes a small number of isolates from man and other animal hosts (cattle, small ruminants and horse), that occurred probably due to spillover infections. With one exception, all isolates from humans were from biovar 1, 3 or 4, which is in agreement with the known virulence of those biovars for man and the belief that biovar 2 is not a zoonotic threat (EFSA, 2009).

Wild boar is an important threat regarding *B. suis* biovar 2 infection, representing an important hazard particularly for the Iberian pig population reared in outdoor breeding systems (Muñoz *et al.*, 2010; EFSA, 2009) but this infection in other animal species should not be ignored. In fact, two isolates from sheep, one from goat and two from bovine were isolated in Portugal and France. We agree with other authors that this situation could become of great concern if brucellosis control programs in domestic pigs are envisaged (Muñoz *et al.*, 2010).

The evolutionary relationships inferred from the MLVA-11 genotypes revealed a *B. suis* population with high genetic divergence among strains based on their host species. The ancestor of the domesticated pig is the wild boar (*Sus scrofa*), which is one of the most numerous and widespread large mammals (Chen *et al.*, 2007). Considering the two biovar 2 lineages circulating in Europe, strains from swine were closely related to strains isolated in both wild boars and hares, supporting the idea that wild animals are a source of brucellosis infection for domestic pigs and also suggesting the movement of the pathogen between regions due to natural dispersal or translocation of wild boars, fact that was already pointed out by other authors (Kreizinger *et al.*, 2014). Likewise, we can speculate that biovar 2 Iberian clonal lineage evolved from the Central-European one and may be the result of an allopatric speciation event, since no isolates of this lineage were obtained above the geographical barriers formed by Ebro River and Pyrenees. MLVA-11 genotyping also disclosed the interconnectedness of *B. suis* biovar 1, 3 and 4 with the Central-European clonal lineage.

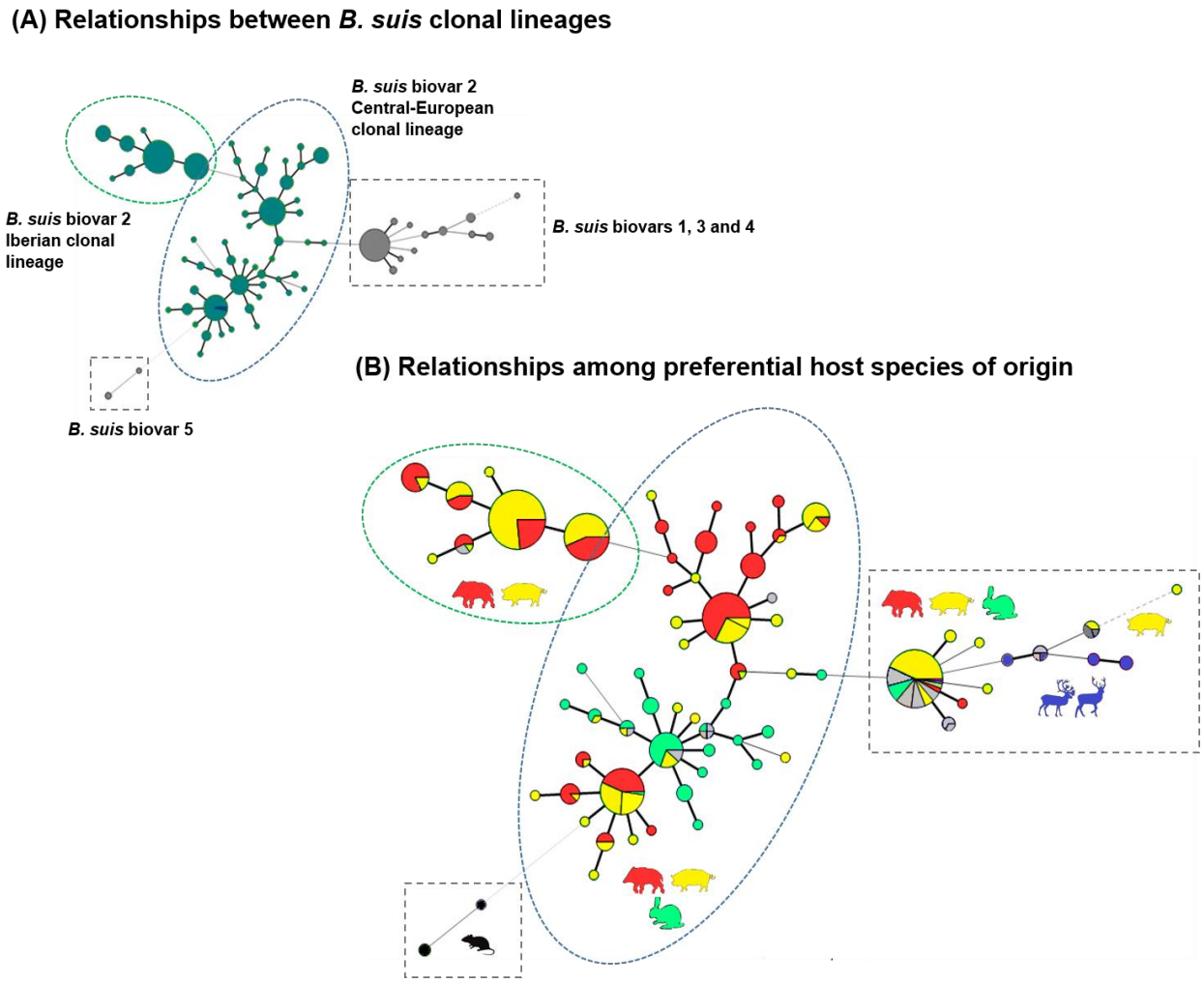


Figure 2.2.2. Minimum spanning tree (MST) analysis of *B. suis* strains using the MLVA-11 typing data. The MST was constructed with a categorical coefficient and the analysis was performed using the typing data from 526 strains, including those from this study and others previously published (Duvnjak *et al.*, 2015; Kreizinger *et al.*, 2014; Jiang *et al.*, 2013; Garcia-Yoldi *et al.*, 2007; Le Flèche *et al.*, 2006) and available in the *Brucella* database hosted at http://microbesgenotyping.i2bc.paris-saclay.fr. The collection includes 85 strains from biovar 1, 422 from biovar 2, four from biovar 3, 11 from biovar 4, three from biovar 5 and one rough strain. (A) MST showing relationships between *B. suis* clonal lineages. Colour code is associated with *B. suis* biovars: biovar 2 strains are shown in blue and all the other biovars are in grey. Each clonal lineage is highlighted by surrounding dotted lines. Two clonal lineages (Iberian and Central-European) are considered for biovar 2 strains, one for biovar 5 strains and another that includes all strains from biovar 1, 3 and 4. The size of the circle indicates the number of strains described in each MLVA-11 genotype and the length of the branches represents the distance between genotypes. (B) Same MST analysis but emphasizing relationships among host species. Strains were marked according to their host species of origin. Only preferential hosts were highlighted: wild boar and swine in biovar 2 Iberian clonal lineage; wild boar, swine and hare in biovar 2 Central-European clonal lineage and biovar 1 strains; swine for biovar 3 strains, and reindeer and caribou for biovar 4. All strains isolated from secondary host species are in grey.

4. Conclusion

The results obtained in this work corroborate that the Suis-ladder multiplex PCR allows a fast and precise identification of *B. suis* at the biovar level while the PCR-RFLP analysis of genes *omp2a* and *omp2b* and gene *omp31* shown to be useful for differentiation between biovars and demonstrated additional polymorphism within biovar 2. MLVA confirmed both a clear distinction between biovars, as well as the close genetic relationship among isolates within the species. While *B. suis* biovars 1, 3 and 4 grouped together in the same cluster (or lineage), biovar 2 were placed into two closely related MLVA clusters according to their geographic origins and PCR-RFLP haplotypes, defining the Iberian (Portugal and Spain) and the Central-European clonal lineages. These results were reinforced when a larger number of strains were included in the MLVA analysis.

In addition, the genomic relationships based on the MLVA-11 genotypes revealed significant associations between *B. suis* biovars and host species, supporting the hypothesis that host adaptation contributes to this structure. Hints were also obtained for the evolution of the Iberian clonal lineage from the Central-European one, probably associated to restrictions of wild boars movement across regions. Nevertheless, a more accurate understanding of the clonal evolution of this pathogen can only be achieved by comparative whole-genome sequence analysis of *B. suis* biovar 2 strains from both clonal lineages.

5. Supplementary material

Supplementary Table S2.2.1. Full information on *B. suis* isolates, including geographic origin and molecular typing data obtained with Suis-ladder multiplex PCR and RFLP-PCR.

Supplementary Table S2.2.2. MLVA-16 allelic profile for the set of 181 *B. suis* strains.

6. References

1. **Aftab, H., Dargis, R., Christensen, J.J., Le Flèche, P. & Kemp, M. (2011).** Imported brucellosis in Denmark: molecular identification and multiple-*locus* variable number tandem repeat analysis (MLVA) genotyping of the bacteria. Scandinavian Journal of Infectious Diseases, 43: 536–8.
2. **Al Dahouk, S., Le Flèche, P., Nöckler, K., Jacques, I., Grayon, M., Scholz, H.C., Tomaso, H., Vergnaud, G. & Neubauer, H. (2007).** Evaluation of *Brucella* MLVA typing for human brucellosis. Journal of Microbiological Methods, 69, 137–145.

3. **Al Dahouk, S., Nöckler. K., Tomaso. H., Splettstoesser, W.D., Jungersen, G., Riber, U., Petry, T., Hoffmann, D., Scholz, H.C., Hensel, A. & Neubauer, H. (2005)**. Seroprevalence of brucellosis, tularemia, and yersiniosis in wild boars (*Sus scrofa*) from north-eastern Germany. Journal of veterinary medicine. B, Infectious diseases and veterinary public health, 52 (10), 444-455.
4. **Allen, A., Breadon, E., Byrne, A., Mallon, T., Skuce, R., Groussaud, P., Dainty, A., Graham, J., Jones, K., Pollock, L. & Whatmore, A. (2015).** Molecular Epidemiology of *Brucella abortus* in Northern Ireland-1991 to 2012. PLoS One, 10(9): e0136721. doi: 10.1371/journal.pone.0136721.
5. **Alton, G.G., Jones, L.M., Angus, R.D. & Verger, J.M. (1988).** Techniques for the Brucellosis Laboratory. Institut National de la Recherche Agronomique, Paris, France.
6. **Chen, K., Baxter, T., Muir, W.M., Groenen, M.A. & Schook, L.B. (2007).** Genetic resources, Genome Mapping and Evolutionary Genomics of the Pig (Sus scrofa). International Journal of Biological Sciences, 3: 153-165.
7. **Cloeckaert, A., Verger, J.M., Grayon, M. & Grépinet, O. (1995).** Restriction site polymorphism of the genes encoding the major 25 and 36 kDa outer-membrane proteins of *Brucella*. Microbiology, 141: 2111–2121.
8. **Cvetnic, Z., Spicic, S., Curic, S., Jukic, B., Lojkic, M., Albert, D., Thiebaud, M. & Garin-Bastuji, B. (2005).** Isolation of *Brucella suis* biovar 3 from horses in Croatia. Veterinary Record, 156: 584–585.
9. **De Massis, F., Ancora, M., Atzeni, M., Rolesu, S., Bandino, E., Danzetta, M.L., Zilli, K., Di Giannatale, E. & Scacchia, M. (2015).** MLVA as an Epidemiological Tool To Trace Back *Brucella melitensis* biovar 1 Re-Emergence in Italy. Transboundary and Emerging Diseases, 62(5): 463-9.
10. **De Massis, F., Di Provvido, A., Di Sabatino, D., Di Francesco, D., Zilli, K., Ancora, M. & Tittarelli, M. (2012).** Isolation of *Brucella suis* biovar 2 from a wild boar in the Abruzzo Region of Italy. Veterinaria Italiana, 48(4):397-404.
11. **Djuricic, B. (2010).** Brucellosis in the Republic of Serbia – The epizootiological Situation. Macedonian Journal of Medical Sciences, 3(3):246-250.
12. **Duvnjak, S., Račić, I., Špičić, S., Zdelar-Tuk, M., Reil, I. & Cvetnić, Ž. (2015).** Characterization of *Brucella suis* isolates from Southeast Europe by multi-*locus* variable-number tandem repeat analysis. Veterinary Microbiology, 180(1-2): 146-50.
13. **EFSA, European Food Safety Authority (2009).** Porcine brucellosis (*Brucella suis*): scientific opinion of the panel on animal health and welfare. EFSA Journal, 1144: 1–112.
14. **Ferrão-Beck, L., Cardoso, R., Muñoz, P.M., de Miguel, M.J., Albert, D., Ferreira, A.C., Marín, C.M., Thiébaud, M., Jacques, I., Grayon, M., Zygmunt, M.S., Garin-Bastuji, B., Blasco, J.M. & Sá, M.I. (2006).** Development of a multiplex PCR assay for polymorphism analysis of *Brucella suis* biovars causing brucellosis in swine. Veterinary Microbiology, 115(1-3): 269-77.
15. **Ferreira, A.C., Almendra, C., Cardoso, R., Silva Pereira, M., Beja-Pereira, A., Luikart, G., Corrêa de Sá, M.I. (2012a)**. Development and evaluation of a selective medium for *Brucella suis*. Research in Veterinary Science, 93: 565–567.
16. **Ferreira, A.C., Chambel, L., Tenreiro, T., Cardoso, R., Flor, L., Dias, I.T., Pacheco, T., Garin-Bastuji, B., Le Flèche, P., Vergnaud, G., Tenreiro, R. & de Sá, M.I. (2012b).** MLVA16 typing of Portuguese human and animal *Brucella melitensis* and *Brucella abortus* isolates. PLoS One, 7: e42514.
17. **García-Yoldi, D., Le Fleche, P., De Miguel, M.J., Muñoz, P.M., Blasco, J.M., Cvetnic, Z., Marín, C.M., Vergnaud, G. & López-Goñi, I. (2007).** Comparison of multiple-*locus* variable-number tandem-repeat analysis with other PCR-based methods for typing *Brucella suis* isolates. Journal of Clinical Microbiology, 45(12): 4070-2.
18. **Garin-Bastuji, B., Hars, J., Calvez, D., Thiebaud, M. & Artois, M. (2000).** Brucellose du porc domestique et du sanglier sauvage due a *Brucella suis* biovar 2 en France. Épidémiologie et Santé Animale, 38 : 1-5.
19. **Garofolo, G., Giannatale, D., De Massis, E., Zilli, F., Ancora, K., Cammà, C., Calistri, P. & Foster, J.T. (2013).** Investigating genetic diversity of *Brucella abortus* and *Brucella melitensis* in Italy with MLVA-16. Infection, Genetics and Evolution, 19: 59-70.

20. **Grégoire, F., Mousset, B., Hanre, D., Michaux, C., Walravens, K. & Linden, A. (2012)**. A serological and bacteriological survey of brucellosis in wild boar (*Sus scrofa*) in Belgium. BMC Veterinary Research, 8: 80–93. doi:10.1186/1746-6148-8-80.
21. **Hubalek, Z., Treml, F., Juricova, Z., Hunady, M., Halouzka, J., Janik, V. & Bill, D. (2002)**. Serological survey of the wild boar (*Sus scrofa*) for tularaemia and brucellosis in South Moravia Veterinary Medicine - Czech,, 47,60–66.
22. **Hunter, P.R. & Gaston, M.A. (1988).** Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. Journal of Clinical Microbiology, 26: 2465–2466.
23. **Jiang, H., Wang, H., Xu, L., Hu, G., Ma, J., Xiao, P., Fan, W., Di, D., Tian, G., Fan, M., Mi, J., Yu, R., Song, L., Zhao, H., Piao, D. & Cui, B. (2013).** MLVA genotyping of *Brucella melitensis* and *Brucella abortus* isolates from different animal species and humans and identification of *Brucella suis* vaccine strain S2 from cattle in China. PLoS One, 8(10):e76332.
24. **Jungersen, G., Sorensen, V., Giese, S.B., Stack, J.A. & Riber, U. (2006)**. Differentiation between serological responses to *Brucella suis* and Yersinia enterocolitica serotype O:9 after natural or experimental infection in pigs. Epidemiology & Infection, 134 (2): 347-357.
25. **Kiliç, S., Ivanov, I.N., Durmaz, R., Bayraktar, M.R. & Ayaslioglu, E. (2011).** Multiple *locus* variable-number tandem-repeat analysis genotyping of human *Brucella* isolates from Turkey. Journal of Clinical Microbiology, 49: 3276–83.
26. **Koppel, C., Knopf, L., Ryser, M.P., Miserez, R., Thur, B. & Stark, K.D.C. (2007)** Serosurveillance for selected infectious disease agents in wild boars (*Sus scrofa*) and outdoor pigs in Switzerland. European Journal of Wildlife Research, 53: 212–220.
27. **Kreizinger, Z., Foster, J.T., Rónai, Z., Sulyok, K.M., Wehmann, E., Jánosi, S. & Gyuranecz, M. (2014).** Genetic relatedness of *Brucella suis* biovar 2 isolates from hares, wild boars and domestic swine. Veterinary Microbiology, 172(3-4): 492-8.
28. **Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoeud, F., Nöckler, N., Neubauer, H., Guilloteau, L.A. & Vergnaud, G. (2006).** Evaluation and selection of tandem repeat *loci* for a *Brucella* MLVA typing assay. BMC Microbiology, 6: 9.
29. **Li, Z.J., Cui, B.Y., Chen, H., Chen, J.D., Zhao, H.Y., Piao, D.R., Jiang, H., Zhang, L., Tang, X., Ke, C.W., Yao, Z. & Tian, G.Z. (2013).** Molecular typing of *Brucella suis* collected from 1960s to 2010s in China by MLVA and PFGE. Biomedical and Environmental Sciences, 26(6): 504-8.
30. **López-Goñi, I., García-Yoldi, D., Marín, C.M., de Miguel, M.J., Barquero-Calvo, E., Guzmán-Verri, C., Albert, D. & Garin-Bastuji, B. (2011).** New Bruce-ladder multiplex PCR assay for the biovar typing of *Brucella suis* and the discrimination of Brucella suis and Brucella canis. Veterinary Microbiology, 154: 152–155.
31. **Muñoz, P.M., Boadella, M., Arnal, M., de Miguel, M.J., Revilla, M., Martínez, D., Vicente, J., Acevedo, P., Oleaga, A., Ruiz-Fons, F., Marín, C.M., Prieto, J.M., de la Fuente, J., Barral, M., Barberán, M., de Luco, D.F., Blasco, J.M. & Gortázar, C. (2010).** Spatial distribution and risk factors of Brucellosis in Iberian wild ungulates. BMC Infectious Diseases, 10: 46.
32. **Scholz, H.C. & Vergnaud, G. (2013).** Molecular characterization of *Brucella* species. Revue Scientifique et Technique-Office International des Epizooties, 32(1): 149-62.
33. **Shevtsov, A., Ramanculov, E., Shevtsova, E., Kairzhanova, A., Tarlykov, P., Filipenko, M., Dymova, M., Abisheva, G., Jailbekova, A., Kamalova, D., Chsherbakov, A., Tulegenov, S., Akhmetova, A., Sytnik, I., Karibaev, T. & Mukanov, K. (2015).** Genetic diversity of *Brucella abortus* and *Brucella melitensis* in Kazakhstan using MLVA-16. Infection, Genetics and Evolution, 34: 173-80.
34. **Sun, M.J., Di, D.D., Li, Y., Zhang, Z.C., Yan, H., Tian, L.L., Jing, Z.G., Li, J.P., Jiang, H. & Fan, W.X. (2016)**. Genotyping of *Brucella melitensis* and *Brucella abortus* strains currently circulating in Xinjiang, China. Infection, Genetics and Evolution, 44: 522-9.
35. **Szulowski, K., Iwaniak, W., Weiner, M. & Złotnicka, J. (2013)**. *Brucella suis* biovar 2 isolations from cattle in Poland. Annals of Agricultural and Environmental Medicine, 20(4): 672-5.

36. **Tay, B.Y., Ahmad, N., Hashim, R., Mohamed, Zahidi, J., Thong, K.L., Koh, X.P. & Mohd Noor, A. (2015).** Multiple-*locus* variable-number tandem-repeat analysis (MLVA) genotyping of human *Brucella* isolates in Malaysia. BMC Infectious Diseases, 15: 220.
37. **Valdezate, S., Navarro, A., Villalón, P., Carrasco, G. & Saéz-Nieto, J.A. (2010).** Epidemiological and phylogenetic analysis of Spanish human *Brucella melitensis* strains by multiple-*locus* variable-number tandem-repeat typing, hypervariable octameric oligonucleotide fingerprinting, and *rpoB* typing. Journal of Clinical Microbiology, 8: 2734–40.
38. **Vizcaíno, N., Verger, J.M., Grayon, M., Zygmunt, M.S. & Cloeckaert, A. (1997).** DNA polymorphism at the *omp31 locus* of *Brucella* spp.: evidence for a large deletion in *Brucella abortus*, and other species-specific markers. Microbiology, 143: 2913–2921.

Supplementary Table S2.2.1. Full information on *B. suis* isolates, including geographic origin and molecular typing data obtained with Suis-ladder multiplex PCR and RFLP-PCR.

Strain No.	Strain Identification	bv	Host[a]	Year	Country (geographic region)	Suis-ladder [b]	RFLP-PCR analysis[c]							
							omp2a ***Nco*I**	*omp2a* ***Sty*I**	*omp2b* ***Eco*RI**	*omp2b* ***Kpn*I**	*omp31* ***Ava*II**	*omp31* ***Hae*III**	Haplotype	Clonal lineage[d]
1	INRA03-26	1	SW	2003	France	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
2	AFSSA-03/3081-2	1	G	2003	Croatia	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
3	AFSSA-04/115	1	H	2004	France	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
4	AFSSA-04/3025-3	1	SW	2004	Croatia	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
5	REF 1330 [biovar 1; ATCC 23444]	1	SW	n.a	United States	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
6	AFSSA-04/2987	1	M	2004	France	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
7	VLA-92/29	1	M	1992	Mexico	S1	P2	P2	P1	P2	P1	P1	1a	n.a.
8	AFSSA-01/5744	1	SW	2001	French Polinesia	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
9	AFSSA-03/2067-203	1	SW	2003	France	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
10	VLA-64/24	1	SW	1964	United States	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
11	VLA-F1/04	1	M	2004	Netherlands	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
12	AFSSA-96/1646-01	1	SW	1996	France	S1	P2	P2	P1	P1	P1	P1	1a	n.a.
13	REF 686 [biovar 3; ATCC 23446]	3	SW	n.a	United States	S3	P2	P2	P1	NC	P1	P1	3a	n.a.
14	VLA-UK6/04	3	M	2004	India	S3	P2	P2	P1	NC	P1	P1	3a	n.a.
15	VLA-63/252	4	C	1963	Alaska	S4	P2	P2	P1	NC	P1	P1	4a	n.a.
16	REF 40 [biovar 4; ATCC 23447]	4	R	n.a	Former USSR	S4	P2	P2	P1	NC	P1	P1	4a	n.a.
17	VLA-63/202	4	R	1963	Poland	S4	P2	P2	P1	NC	P1	P1	4a	n.a.
18	VLA-63/219	4	R	1963	Poland	S4	P2	P2	P1	NC	P1	P1	4a	n.a.
19	VLA-63/198	4	R	1963	Poland	S4	P2	P2	P1	NC	P1	P1	4a	n.a.
20	AFSSA-92/11580-4528	2	H	1992	France	S2	NC	NC	P3	NC	P2	P1	2b	CE
21	AFSSA-96/9635	2	SW	1996	France	S2	NC	NC	P1	P1	P2	P2	2c	CE
22	A183	2	WB	n.a	Germany	S2	NC	NC	P3	NC	P2	P2	2a	CE
23	A196	2	WB	n.a	Germany	S2	NC	NC	P3	NC	P2	P2	2a	CE
24	04RB0377	2	WB	n.a	Germany	S2	NC	NC	P3	NC	P2	P2	2a	CE
25	05RB0007	2	WB	n.a	Germany	S2	NC	NC	P3	NC	P2	P2	2a	CE
26	AFSSA-04/3025-1	2	SW	2004	Croatia	S2	NC	NC	P1	P1	P2	P2	2c	CE
27	05RB1442	2	H	n.a	Germany	S2	NC	NC	P3	NC	P2	P2	2a	CE

Strain No.	Strain Identification	bv	Host[a]	Year	Country (geographic region)	Suis-ladder [b]	RFLP-PCR analysis[c]							
							omp2a ***Nco*I**	*omp2a* ***Sty*I**	*omp2b* ***Eco*RI**	*omp2b* ***Kpn*I**	*omp31* ***Ava*II**	*omp31* ***Hae*III**	Haplotype	Clonal lineage[d]
28	REF Thomsen [biovar 2; ATCC 23445]	2	SW	1952	Denmark	S2	NC	NC	P3	NC	P2	P2	2a	CE
29	AFSSA-92/13000	2	H	1992	France	S2	NC	NC	P3	NC	P2	P1	2b	CE
30	VLA-74/11	2	n.a	1974	Denmark	S2	NC	NC	P3	NC	P2	P2	2a	CE
31	AFSSA-00/9182	2	H	2000	France	S2	NC	NC	P3	NC	P2	P1	2b	CE
32	Bs364CITA	2	WB	2008	Spain (Añón)	S2	NC	NC	P1	P1	P2	P2	2c	CE
33	Bs365CITA	2	WB	2008	Spain (Añón)	S2	NC	NC	P1	P1	P2	P2	2c	CE
34	Bs396CITA	2	WB	2009	Spain (Navarra)	S2	NC	NC	P1	P1	P2	P2	2c	CE
35	AFSSA-04/1918-1	2	WB	2004	Switzerland	S2	NC	NC	P1	P1	P2	P2	2c	CE
36	C13B4	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
37	C2B11	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
38	C9B3	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
39	PY69	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
40	0111602/4+9	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
41	0111602/3+8	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
42	It4	2	WB	n.a	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
43	It5	2	WB	n.a	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
44	S275	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
45	C9B4	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
46	C8B3	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
47	It2	2	WB	n.a	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
48	It3	2	WB	n.a	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
49	AFSSA-04/770	2	WB	2004	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
50	It6	2	WB	n.a	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
51	It1	2	WB	n.a	Italy	S2	NC	NC	P1	P1	P2	P2	2c	CE
52	AFSSA-97/9757	2	SW	1997	France	S2	NC	NC	P1	P1	P2	P2	2c	CE
53	MASA07	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
54	AFSSA-97/4924-10	2	SW	1997	France	S2	NC	NC	P1	P1	P2	P2	2c	CE
55	AFSSA-03/1483-8	2	WB	2003	France	S2	NC	NC	P1	P1	P2	P2	2c	CE
56	C13B1	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
57	C3B3	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
58	C13B6	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE

Strain No.	Strain Identification	bv	Host[a]	Year	Country (geographic region)	Suis-ladder [b]	RFLP-PCR analysis[c]							
							omp2a *Nco*I	*omp2a* *Sty*I	*omp2b* *Eco*RI	*omp2b* *Kpn*I	*omp31* *Ava*II	*omp31* *Hae*III	Haplotype	Clonal lineage[d]
59	C11B4	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
60	C4B6	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
61	S120	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
62	C5B5	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
63	COSA13	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
64	RATES5-11	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
65	C6B1	2	WB	n.a	Belgium	S2	NC	NC	P1	P1	P2	P2	2c	CE
66	AFSSA-00/4898	2	G	2000	France	S2	NC	NC	P1	P1	P2	P2	2c	CE
67	AFSSA-98/7296-4204	2	H	1998	France	S2	NC	NC	P3	NC	P2	P2	2a	CE
68	AFSSA-98/6335	2	SW	1998	France	S2	NC	NC	P3	NC	P2	P1	2b	CE
69	LNIV-2948(3)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
70	LNIV-2948(10)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
71	LNIV-2948(12)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
72	LNIV-2948(20)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
73	LNIV-2948(27)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
74	LNIV-2948(29)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
75	LNIV-2948(34)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
76	Bs145	2	SW	n.a	Spain	S2	P2	P2	P1	P1	P2	P2	2e	IB
77	S-145(PN-II)	2	WB	2005	Spain (Asturias)	S2	P2	P2	P1	P1	P2	P2	2e	IB
78	Bs144	2	SW	n.a	Spain	S2	P2	P2	P1	P1	P2	P2	2e	IB
79	Bs143CITA	2	WB	2005	Spain (Asturias)	S2	P2	P2	P1	P1	P2	P2	2e	IB
80	PT09172	2	WB	2009	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
81	LNIV-44821(121)/09	2	WB	2009	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
82	LNIV-44821(122)/09	2	WB	2009	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
83	LNIV-44821(123)/09	2	WB	2009	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
84	LNIV-46685(20)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
85	S22	2	SW	2000	Spain (Salamanca)	S2	P2	P2	P3	NC	P2	P2	2d	IB
86	Bs147	2	SW	n.a	Spain	S2	P2	P2	P3	NC	P2	P2	2d	IB
87	Bs146	2	SW	n.a	Spain	S2	P2	P2	P3	NC	P2	P2	2d	IB
88	S4	2	SW	1992	Spain (Salamanca)	S2	P2	P2	P3	NC	P2	P2	2d	IB
89	S21	2	SW	1999	Spain (Huelva)	S2	P2	P2	P3	NC	P2	P2	2d	IB

Strain No.	Strain Identification	bv	Host[a]	Year	Country (geographic region)	Suis-ladder [b]	RFLP-PCR analysis[c]						Haplotype	Clonal lineage[d]
							omp2a *Nco*I	*omp2a* *Sty*I	*omp2b* *Eco*RI	*omp2b* *Kpn*I	*omp31* *Ava*II	*omp31* *Hae*III		
90	S6	2	SW	1998	Spain (Badajoz)	S2	P2	P2	P3	NC	P2	P2	2d	IB
91	S1	2	SW	1992	Spain	S2	P2	P2	P3	NC	P2	P2	2d	IB
92	S2	2	SW	1992	Spain (Salamanca)	S2	P2	P2	P3	NC	P2	P2	2d	IB
93	S3	2	SW	1992	Spain (Salamanca)	S2	P2	P2	P3	NC	P2	P2	2d	IB
94	LNIV-4498(J1)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
95	LNIV-4498(J9)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
96	LNIV-4498(J4)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
97	LNIV-4498(J6)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
98	VLA-63/311	2	SW	1963	Unknown	S2	P2	P2	P3	NC	P2	P2	2d	IB
99	LNIV-9789(J9)/08	2	WB	2008	Portugal (Entre Douro e Minho)	S2	P2	P2	P3	NC	P2	P2	2d	IB
100	LNIV-19122(J1)/08	2	WB	2008	Portugal (Entre Douro e Minho)	S2	P2	P2	P3	NC	P2	P2	2d	IB
101	LNIV-17888(J3)/08	2	WB	2008	Portugal (Entre Douro e Minho)	S2	P2	P2	P3	NC	P2	P2	2d	IB
102	LNIV-21346(J2)/08	2	WB	2008	Portugal (Entre Douro e Minho)	S2	P2	P2	P3	NC	P2	P2	2d	IB
103	LNIV-2454(J10)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
104	LNIV-1344(2)/09	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
105	LNIV-2948(1)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
106	LNIV-2948(5)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
107	PT09143	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
108	LNIV-2948(8)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
109	LNIV-2948(9)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
110	LNIV-4189(1)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
111	LNIV-4189(2)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
112	LNIV-8605(2)/10	2	WB	2010	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
113	LNIV-8605(9)/10	2	WB	2010	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
114	S12	2	SW	1992	Spain (Salamanca)	S2	P2	P2	P1	P1	P2	P2	2e	IB
115	LNIV-4647(1)/09	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
116	LNIV-45014(4)/08	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P1	P1	P2	P2	2e	IB
117	LNIV-4477(4)/09	2	WB	2009	Portugal (Algarve)	S2	P2	P2	P1	P1	P2	P2	2e	IB
118	LNIV-J2A/08	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P1	P1	P2	P2	2e	IB
119	S27	2	SW	2000	Spain (Badajoz)	S2	P2	P2	P1	P1	P2	P2	2e	IB
120	S13	2	SW	1992	Spain (Salamanca)	S2	P2	P2	P1	P1	P2	P2	2e	IB

Strain No.	Strain Identification	bv	Host[a]	Year	Country (geographic region)	Suis-ladder [b]	RFLP-PCR analysis[c]							
							omp2a NcoI	***omp2a StyI***	***omp2b EcoRI***	***omp2b KpnI***	***omp31 AvaII***	***omp31 HaeIII***	**Haplotype**	**Clonal lineage[d]**
121	LNIV-44821(81)/09	2	WB	2009	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
122	LNIV-44821(86)/09	2	WB	2009	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
123	LNIV-44821(79)/09	2	WB	2009	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
124	LNIV-3115(9)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
125	LNIV-3514(2)/10	2	WB	2010	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
126	LNIV-3515(6)/10	2	WB	2010	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
127	LNIV-44401(3)/08	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
128	LNIV-Soc57/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
129	LNIV-Soc64/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
130	LNIV-Soc73/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
131	LNIV-Soc41/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
132	LNIV-1344(3)/09	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
133	LNIV-14256(72)/09	2	SW	2009	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
134	LNIV-231(1)/09	2	SW	2009	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
135	LNIV-14256(40)/09	2	SW	2009	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
136	LNIV-1989(112)/10	2	WB	2010	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
137	LNIV-3115(13)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
138	LNIV-4478(4)/09	2	WB	2009	Portugal (Algarve)	S2	P2	P2	P3	NC	P2	P2	2d	IB
139	LNIV-4187(7)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
140	LNIV-3478(3)/09	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
141	LNIV-438/00	2	SW	2000	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
142	LNIV-5346(1)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
143	LNIV-5346(16)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
144	LNIV-5346(3)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
145	LNIV-6552(3)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
146	LNIV-6552(8)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
147	LNIV-6552(2)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
148	LNIV-2739(18)/02	2	SW	2002	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
149	LNIV-2739(9)/02	2	SW	2002	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
150	AFSSA-PT-6552/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
151	LNIV-4215(20)/01	2	SW	2001	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB

Strain No.	Strain Identification	bv	Host[a]	Year	Country (geographic region)	Suis-ladder [b]	RFLP-PCR analysis[c]							
							omp2a *Nco*I	*omp2a* *Sty*I	*omp2b* *Eco*RI	*omp2b* *Kpn*I	*omp31* *Ava*II	*omp31* *Hae*III	Haplotype	Clonal lineage[d]
152	LNIV-6552(10)/00	2	SW	2000	Portugal (Ribatejo e Oeste)	S2	P2	P2	P3	NC	P2	P2	2d	IB
153	LNIV-2739(3)/02	2	SW	2002	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
154	LNIV-468/00	2	SW	2000	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
155	LNIV-Soc29/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
156	LNIV-Soc43/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
157	LNIV-7676/00	2	B	2000	Portugal (Algarve)	S2	P2	P2	P3	NC	P2	P2	2d	IB
158	LNIV-1262/00	2	SW	2000	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
159	LNIV-Soc39/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
160	LNIV-Soc35/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
161	LNIV-Soc50/00	2	SW	2000	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
162	LNIV-1967/00	2	SW	2000	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
163	LNIV-46685(15)/08	2	WB	2008	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
164	LNIV-9789(J10)/08	2	WB	2008	Portugal (Entre Douro e Minho)	S2	P2	P2	P3	NC	P2	P2	2d	IB
165	LNIV-9789(Feto)/08	2	WB	2008	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
166	LNIV-44406(3)/08	2	WB	2009	Portugal (Beira Interior)	S2	P2	P2	P3	NC	P2	P2	2d	IB
167	S32	2	SW	2000	Spain (Toledo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
168	LNIV-5414(12)/03	2	SW	2003	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
169	LNIV-5414(13)/03	2	SW	2003	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
170	LNIV-5414(2)/03	2	SW	2003	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
171	LNIV-5414(13)/03	2	SW	2003	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
172	S34	2	SW	2000	Spain (Toledo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
173	LNIV-21566(94)/09	2	WB	2009	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
174	LNIV-S-384(679)	2	SW	2009	Spain (Piensos del segre)	S2	P2	P2	P3	NC	P2	P2	2d	IB
175	LNIV-22497(475)/08	2	S	2008	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
176	LNIV-22498(697)/08	2	S	2008	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
177	LNIV-4193(1)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
178	LNIV-4193(2)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
179	LNIV-4193(9)/09	2	WB	2009	Portugal (Alentejo)	S2	P2	P2	P3	NC	P2	P2	2d	IB
180	LNIV-9H	2	SW	2011	Portugal (Trás-os-Montes)	S2	P2	P2	P3	NC	P2	P2	2d	IB
181	REF 513 [biovar 5; NCTC 11996]	5	WR	n.a	Former USSR	S5	NC	NC	P1	P1	P1	P1	5a	n.a.

[a] Host: B, bovine; C, caribou; G, goat; H, hare; M, man; R, reindeer; S, sheep; S, swine; WB, wild boar; WR, wild rodent; n.a., information not available.

[b] Suis-ladder multiplex PCR according to López-Goñi *et al.*, 2011. S1 to S5 refer to electrophoresis profile observed for each *B. suis* biovar.

[c] PCR-RFLP analysis *of omp31, omp2a* and *omp2b* genes using the restriction enzymes AvaII*, HaeIII, NcoI, StyI, EcoRI* and *KpnI*. P1 to P3 refer to restriction patterns observed for each enzyme; NC: not cleaved.

[d] Clonal lineage of biovar 2 strains: CE, Central-European; IB: Iberian; n.a.: not applicable.

Supplementary Table S2.2.2. MLVA-16 allelic profile for the set of 181 *B. suis* strains.

Strain no.	Haplotype	Clonal lineage[a]	MLVA-16 molecular markers[b]															MLVA-8[c]	MLVA-11[d]	MLVA-16[e]
			B06	B08	B11	B12	B42	B43	B45	B55	B18	B19	B21	B04	B07	B09	B16	B30			
1	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	5	6	5	5	3	6	33	1
2	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	5	6	5	5	3	6	33	1
3	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	5	6	5	5	3	6	33	1
4	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	5	6	5	5	3	6	33	1
5	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	6	6	5	5	3	6	33	2
6	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	6	6	5	5	3	6	33	2
7	1a	n.a.	2	3	6	10	4	1	5	2	5	38	9	6	6	6	5	3	6	32	3
8	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	7	7	5	5	4	6	33	4
9	1a	n.a.	2	3	6	10	4	1	5	2	6	43	9	5	6	5	5	4	6	256	5
10	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	5	5	7	4	3	6	33	6
11	1a	n.a.	2	3	6	10	4	1	5	2	4	38	9	5	4	10	6	3	6	33	7
12	1a	n.a.	2	4	6	10	4	1	5	2	6	38	9	7	7	6	5	3	153	255	8
13	3a	n.a.	2	3	4	11	3	1	5	2	4	38	9	8	6	10	4	5	4	31	9
14	3a	n.a.	2	3	4	11	3	1	5	2	4	38	9	7	5	12	4	7	4	31	10
15	4a	n.a.	2	3	9	10	3	1	5	2	4	40	9	5	6	11	9	5	5	30	11
16	4a	n.a.	2	3	9	11	3	1	5	2	5	36	9	4	5	9	6	3	3	27	12
17	4a	n.a.	2	3	9	10	3	1	5	2	5	36	9	4	5	9	6	3	5	260	13
18	4a	n.a.	2	3	9	10	3	1	5	2	5	36	9	8	5	10	9	3	5	260	14
19	4a	n.a.	2	3	9	10	3	1	5	2	5	36	9	12	8	16	5	3	5	260	15
20	2b	CE	2	3	8	14	6	1	5	2	4	38	9	11	5	21	2	4	162	254	16
21	2c	CE	2	3	8	14	6	1	5	7	4	38	9	23	7		2	5	163	253	17
22	2a	CE	2	5	8	14	6	1	5	7	4	43	9	6	14	20	2	6	158	237	18
23	2a	CE	2	5	8	14	6	1	5	7	4	43	9	6	14	19	2	6	158	237	19
24	2a	CE	2	5	8	14	6	1	5	7	4	43	9	6	14	18	2	6	158	237	20
25	2a	CE	2	5	8	14	6	1	5	7	4	43	9	8	9	10	2	6	158	237	21
26	2c	CE	2	5	8	14	6	1	5	7	4	43	9	12	5	3	2	8	158	237	22
27	2a	CE	2	5	8	14	6	1	5	7	6	43	9	12	12	16	2	6	158	236	23
28	2a	CE	2	4	8	14	6	1	5	2	6	43	9	9	9	18	2	4	9	50	24
29	2b	CE	2	4	8	14	6	1	5	2	6	43	9	14	12	5	2	5	9	50	25
30	2a	CE	2	4	8	14	6	1	5	2	6	43	9	20	9	10	2	4	9	50	26

Strain no.	Haplotype	Clonal lineage[a]	MLVA-16 molecular markers[b]															MLVA-8[c]	MLVA-11[d]	MLVA-16[e]
			B06	**B08**	**B11**	**B12**	**B42**	**B43**	**B45**	**B55**	**B18**	**B19**	**B21**	**B04**	**B07**	**B09**	**B16**	**B30**			
31	2b	CE	2	4	8	14	6	1	5	2	6	43	9	13	5	9	2	4	9	50	27
32	2c	CE	2	5	8	14	6	1	5	6	6	43	9	16	7	18	2	11	157	244	28
33	2c	CE	2	5	8	14	6	1	5	6	6	43	9	14	7	18	2	10	157	244	29
34	2c	CE	2	5	8	14	6	1	5	6	6	43	9	23	9	7	2	10	157	244	30
35	2c	CE	2	4	8	14	6	1	5	6	6	43	9	20	7	12	2	9	7	44	31
36	2c	CE	2	4	8	14	6	1	5	6	6	43	9	22	7	17	2	9	7	44	32
37	2c	CE	2	4	8	14	6	1	5	6	6	43	9	15	9	17	2	9	7	44	33
38	2c	CE	2	4	8	14	6	1	5	6	6	43	9	15	9	18	2	9	7	44	34
39	2c	CE	2	4	8	14	6	1	5	6	6	43	9	15	7	15	2	9	7	44	35
40	2c	CE	2	4	8	14	6	1	5	6	4	43	9	15	7	7	2	9	7	41	36
41	2c	CE	2	4	8	14	6	1	5	6	4	43	9	15	7	7	2	9	7	41	36
42	2c	CE	2	4	8	14	6	1	5	6	6	43	9	7	6	18	2	9	7	44	37
43	2c	CE	2	4	8	14	6	1	5	6	6	43	9	7	6	18	2	9	7	44	37
44	2c	CE	2	4	8	14	6	1	5	6	6	43	9	7	5	15	2	9	7	44	38
45	2c	CE	2	4	8	14	6	1	5	6	6	43	9	9	5	9	2	9	7	44	39
46	2c	CE	2	4	8	14	6	1	5	6	6	43	9	12	9	15	2	8	7	44	40
47	2c	CE	2	4	8	14	6	1	5	6	6	43	9	7	6	16	2	10	7	44	41
48	2c	CE	2	4	8	14	6	1	5	6	6	43	9	7	6	16	2	10	7	44	41
49	2c	CE	2	4	8	14	6	1	5	6	6	43	9	6	6	16	2	10	7	44	42
50	2c	CE	2	4	8	14	6	1	5	6	6	43	9	8	6	19	2	10	7	44	43
51	2c	CE	2	4	8	14	6	1	5	6	6	43	9	8	6	10	2	11	7	44	44
52	2c	CE	2	4	8	14	6	1	5	6	6	43	9	15	6	7	2	5	7	44	45
53	2c	CE	2	4	8	14	6	1	5	6	6	43	9	11	6	17	2	12	7	44	46
54	2c	CE	2	5	8	14	6	1	5	6	6	43	9	13	6		2	8	157	244	47
55	2c	CE	2	4	8	14	6	1	5	6	4	43	9	9	7	9	2	8	7	41	48
56	2c	CE	2	4	8	14	6	1	5	6	5	43	9	13	8	17	2	10	7	246	49
57	2c	CE	2	4	8	14	6	1	5	6	5	43	9	15	8	7	2	10	7	246	50
58	2c	CE	2	4	8	14	6	1	5	6	5	43	9	14	8	16	2	9	7	246	51
59	2c	CE	2	4	8	14	6	1	5	6	5	43	9	14	8	16	2	9	7	246	51
60	2c	CE	2	4	8	14	6	1	5	6	5	43	9	10	5	19	2	9	7	246	52
61	2c	CE	2	4	8	14	6	1	5	6	5	43	9	15	7	10	2	11	7	246	53
62	2c	CE	2	4	8	14	6	1	5	6	5	43	9	9	7	7	2	11	7	246	54
63	2c	CE	2	4	8	14	6	1	5	6	6	43	9	14	8	11	2	10	7	44	55

Strain no.	Haplotype	Clonal lineage[a]	MLVA-16 molecular markers[b]															MLVA-8[c]	MLVA-11[d]	MLVA-16[e]
			B06	**B08**	**B11**	**B12**	**B42**	**B43**	**B45**	**B55**	**B18**	**B19**	**B21**	**B04**	**B07**	**B09**	**B16**	**B30**			
64	2c	CE	2	4	8	14	6	1	5	6	6	43	9	9	7	16	2	7	7	44	56
65	2c	CE	2	4	8	14	6	1	5	6	6	43	9	11	10	22	2	7	7	44	57
66	2c	CE	2	4	8	14	6	1	5	7	6	43	9	4	7	3	2	7	8	45	58
67	2a	CE	2	4	9	14	6	1	5	7	6	43	9	14	10	12	2	11	15	239	59
68	2b	CE	2	4	8	14	6	1	5	2	7	43	9	4	7	3	2	4	9	48	60
69	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	13	11	2	8	18	61	61
70	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	13	11	2	8	18	61	61
71	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	13	11	2	8	18	61	61
72	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	13	11	2	8	18	61	61
73	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	13	11	2	8	18	61	61
74	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	13	11	2	8	18	61	61
75	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	10	11	2	8	18	61	62
76	2e	IB	2	5	8	9	5	1	5	4	5	38	9	12	5	17	2	8	18	61	63
77	2e	IB	2	5	8	9	5	1	5	5	6	38	9	7	11	10	2	8	19	58	64
78	2e	IB	2	5	8	9	5	1	5	4	5	38	9	7	11	10	2	8	18	61	65
79	2e	IB	2	5	8	9	5	1	5	5	5	38	9	7	12	9	2	8	19	60	66
80	2d	IB	2	5	8	9	5	1	5	5	5	38	9	5	7	12	2	8	19	60	67
81	2d	IB	2	5	8	9	5	1	5	5	5	38	9	5	7	12	2	8	19	60	67
82	2d	IB	2	5	8	9	5	1	5	5	5	38	9	5	7	12	2	8	19	60	67
83	2d	IB	2	5	8	9	5	1	5	5	5	38	9	5	7	12	2	8	19	60	67
84	2d	IB	2	5	8	9	5	1	5	5	5	38	9	14	8	14	2	8	19	60	68
85	2d	IB	2	5	8	9	5	1	5	5	5	38	9	9	7	13	2	7	19	60	69
86	2d	IB	2	5	8	9	5	1	5	5	5	38	9	21	7	11	2	7	19	60	70
87	2d	IB	2	5	8	9	5	1	5	5	5	38	9	20	7	10	2	7	19	60	71
88	2d	IB	2	5	8	9	5	1	5	5	5	38	9	15	7	14	2	7	19	60	72
89	2d	IB	2	5	8	9	5	1	5	5	5	38	9	12	9	13	2	7	19	60	73
90	2d	IB	2	5	8	9	5	1	5	5	5	38	9	12	9	7	2	7	19	60	74
91	2d	IB	2	5	8	9	5	1	5	5	5	38	9	15	9	13	2	7	19	60	75
92	2d	IB	2	5	8	9	5	1	5	5	5	38	9	15	9	13	2	7	19	60	75
93	2d	IB	2	5	8	9	5	1	5	5	5	38	9	15	9	14	2	7	19	60	76
94	2d	IB	2	5	8	9	5	1	5	5	5	38	9	20	9	14	2	7	19	60	77
95	2d	IB	2	5	8	9	5	1	5	5	5	38	9	20	9	14	2	7	19	60	77
96	2d	IB	2	5	8	9	5	1	5	5	5	38	9	20	9	14	2	7	19	60	77

Strain no.	Haplotype	Clonal lineage[a]	MLVA-16 molecular markers[b]															MLVA-8[c]	MLVA-11[d]	MLVA-16[e]
			B06	**B08**	**B11**	**B12**	**B42**	**B43**	**B45**	**B55**	**B18**	**B19**	**B21**	**B04**	**B07**	**B09**	**B16**	**B30**			
97	2d	IB	2	5	8	9	5	1	5	5	5	38	9	20	9	14	2	7	19	60	77
98	2d	IB	2	5	8	9	5	1	5	5	5	38	9	15	9	17	2	8	19	60	78
99	2d	IB	2	5	8	9	5	1	5	5	5	38	9	13	10	17	2	7	19	60	79
100	2d	IB	2	5	8	9	5	1	5	5	5	38	9	13	10	17	2	7	19	60	79
101	2d	IB	2	5	8	9	5	1	5	5	5	38	9	13	10	17	2	7	19	60	79
102	2d	IB	2	5	8	9	5	1	5	5	5	38	9	13	10	17	2	7	19	60	79
103	2d	IB	2	5	8	9	5	1	5	5	5	38	9	13	10	9	2	6	19	60	80
104	2d	IB	2	5	8	9	5	1	5	5	5	38	9	11	6	9	2	7	19	60	81
105	2e	IB	2	5	8	9	5	1	5	4	5	38	9	11	10	9	2	7	18	61	82
106	2e	IB	2	5	8	9	5	1	5	4	5	38	9	11	10	9	2	7	18	61	82
107	2e	IB	2	5	8	9	5	1	5	4	5	38	9	11	10	9	2	7	18	61	82
108	2e	IB	2	5	8	9	5	1	5	4	5	38	9	11	10	9	2	7	18	61	82
109	2e	IB	2	5	8	9	5	1	5	4	5	38	9	11	10	9	2	7	18	61	82
110	2e	IB	2	5	8	9	5	1	5	5	5	38	9	11	10	7	2	4	19	60	83
111	2e	IB	2	5	8	9	5	1	5	5	5	38	9	11	10	7	2	4	19	60	83
112	2e	IB	2	5	8	9	5	1	5	4	5	38	9	12	9	7	2	7	18	61	84
113	2e	IB	2	5	8	9	5	1	5	4	5	38	9	13	9	7	2	7	18	61	85
114	2e	IB	2	5	8	9	5	1	5	4	6	38	9	7	7	7	2	7	18	59	86
115	2d	IB	2	5	8	9	5	1	5	5	5	38	9	9	15	16	2	10	19	60	87
116	2e	IB	2	5	8	9	5	1	5	4	6	38	9	9	12	11	2	7	18	59	88
117	2e	IB	2	5	8	9	5	1	5	4	6	38	9	9	14	15	2	8	18	59	89
118	2e	IB	2	5	8	9	5	1	5	4	6	38	9	13	15	8	2	4	18	59	90
119	2e	IB	2	5	8	9	5	1	5	4	6	38	9	20	9	13	2	5	18	59	91
120	2e	IB	2	5	8	9	5	1	5	4	6	38	9	20	9	13	2	5	18	59	91
121	2d	IB	2	5	8	9	5	1	5	5	6	38	9	13	6	10	2	5	19	58	92
122	2d	IB	2	5	8	9	5	1	5	5	6	38	9	13	6	10	2	5	19	58	92
123	2d	IB	2	5	8	9	5	1	5	5	6	38	9	13	6	17	2	5	19	58	93
124	2d	IB	2	5	8	9	5	1	5	5	6	38	9	10	6	7	2	5	19	58	94
125	2d	IB	2	5	8	9	5	1	5	5	6	38	9	11	6	19	2	9	19	58	95
126	2d	IB	2	5	8	9	5	1	5	5	6	38	9	12	6	19	2	9	19	58	96
127	2d	IB	2	5	8	9	5	1	5	5	6	38	9	20	6	19	2	6	19	58	97
128	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	6	18	2	7	19	58	98
129	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	6	18	2	7	19	58	98

Strain no.	Haplotype	Clonal lineage[a]	MLVA-16 molecular markers[b]															MLVA-8[c]	MLVA-11[d]	MLVA-16[e]
			B06	**B08**	**B11**	**B12**	**B42**	**B43**	**B45**	**B55**	**B18**	**B19**	**B21**	**B04**	**B07**	**B09**	**B16**	**B30**			
130	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	6	18	2	7	19	58	98
131	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	6	18	2	7	19	58	98
132	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	6	10	2	6	19	58	99
133	2d	IB	2	5	8	9	5	1	5	5	6	38	9	21	6	15	2	7	19	58	100
134	2d	IB	2	5	8	9	5	1	5	5	6	38	9	21	6	15	2	7	19	58	100
135	2d	IB	2	5	8	9	5	1	5	5	5	38	9	21	6	16	2	7	19	60	101
136	2d	IB	2	5	8	9	5	1	5	5	6	38	9	3	8	12	2	7	19	58	102
137	2d	IB	2	5	8	9	5	1	5	5	6	38	9	10	5	7	2	6	19	58	103
138	2d	IB	2	5	8	9	5	1	5	5	6	38	9	21	5	7	2	6	19	58	104
139	2d	IB	2	5	8	9	5	1	5	5	6	38	9	21	5	8	2	6	19	58	105
140	2d	IB	2	5	8	9	5	1	5	5	6	38	9	13	5	16	2	6	19	58	106
141	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	107
142	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	107
143	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
144	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
145	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
146	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
147	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
148	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
149	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
150	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
151	2d	IB	2	5	8	9	5	1	5	5	6	38	9	8	5	15	2	6	19	58	108
152	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	15	2	6	19	58	108
153	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	15	2	6	19	58	108
154	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	19	2	6	19	58	109
155	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	19	2	6	19	58	109
156	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	18	2	6	19	58	110
157	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	14	2	6	19	58	111
158	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	17	2	7	19	58	112
159	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	17	2	7	19	58	112
160	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	17	2	6	19	58	113
161	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	17	2	6	19	58	113
162	2d	IB	2	5	8	9	5	1	5	5	6	38	9	9	5	15	2	7	19	58	114

Strain no.	Haplotype	Clonal lineage[a]	MLVA-16 molecular markers[b]															MLVA-8[c]	MLVA-11[d]	MLVA-16[e]
			B06	**B08**	**B11**	**B12**	**B42**	**B43**	**B45**	**B55**	**B18**	**B19**	**B21**	**B04**	**B07**	**B09**	**B16**	**B30**			
163	2d	IB	2	5	8	9	5	1	5	5	7	38	9	9	5	21	2	6	19	233	115
164	2d	IB	2	5	8	9	5	1	5	5	6	38	9	13	5	7	2	8	19	58	116
165	2d	IB	2	5	8	9	5	1	5	5	6	38	9	12	5	7	2	8	19	58	117
166	2d	IB	2	5	8	9	5	1	5	5	6	38	9	17	5	17	2	8	19	58	118
167	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	5	20	2	8	19	58	119
168	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	5	20	2	8	19	58	119
169	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	5	20	2	8	19	58	119
170	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	5	20	2	8	19	58	119
171	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	5	20	2	8	19	58	119
172	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	5	12	2	8	19	58	120
173	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	7	15	2	9	19	58	121
174	2d	IB	2	5	8	9	5	1	5	5	6	38	9	15	8	15	2	9	19	58	122
175	2d	IB	2	5	8	9	5	1	5	5	7	38	9	15	6	14	2	8	19	233	123
176	2d	IB	2	5	8	9	5	1	5	5	7	38	9	15	6	14	2	8	19	233	123
177	2d	IB	2	5	8	9	5	1	5	5	7	38	9	15	5	18	2	7	19	233	124
178	2d	IB	2	5	8	9	5	1	5	5	7	38	9	15	5	18	2	7	19	233	124
179	2d	IB	2	5	8	9	5	1	5	5	7	38	9	15	5	18	2	7	19	233	124
180	2d	IB	2	5	8	9	5	1	5	5	7	38	9	12	4	19	2	6	19	233	125
181	5a	n.a.	1	2	9	14	1	2	5	5	7	44	9	9	5	3	9	5	21	2	126

[a] Clonal lineage of biovar 2 strains: CE, Central-European; IB: Iberian; n.a.: not applicable.

[b] MLVA assay using the 16 markers acording to Ferreira *et al.*, 2012; B, bruce.

[c] MLVA-8 genotypes corresponding to each isolate in the *Brucella* MLVA database (including only minisatellite markers from panels 1).

[d] MLVA-11 genotypes to each isolate in the *Brucella* MLVA database (including mini-/microsatellite markers from panels 1 and 2A).

[e] MLVA-16 genotypes (including all markers).

Chapter 3

Whole-genome mapping reveals a large genetic inversion on Iberian Brucella suis biovar 2 strains

The results presented in this subchapter were previously published.

Ferreira, A.C., Tenreiro, R., Dias, R. & Corrêa de Sá, M.I. (2016). Whole-genome mapping reveals a large chromosomal inversion on Iberian *Brucella suis* biovar 2 strains. *Veterinary Microbiology,* 192: 220–225.

Authors contributions:

ACF developed the PCR protocols and performed all the molecular tests.

ACF and **RD** analyzed the data.

ACF and MICS draft the manuscript.

RT and **RD** revised the manuscript.

1. Introduction

Optical mapping is a recent genomic technology, currently referred to as whole-genome mapping (WGM), which produces ordered genome-wide high-resolution restriction maps (optical maps) from single stained DNA molecules immobilized and digested on an open glass surface and visualized by fluorescence microscopy. By mapping the restriction sites along the DNA, the resulting spectrum serves as a unique fingerprint or barcode for that sequence (Anantharaman and Mishra, 2013; Zhou *et al.*, 2007) and WGM provides higher resolution for epidemiological purposes compared to the gold standard of pulsed-field gel electrophoresis (PFGE) restriction analysis (Miller *et al.*, 2013). With hundreds of markers across the genome, optical maps differentiate strains and precisely describe genomic rearrangements, such as insertions, deletions, duplications, and inversions. WGM has been widely used in microbial genomic studies, with applications in areas of comparative genomics and strain typing (Bosch *et al.*, 2013; Schwan *et al.*, 2010; Shukla *et al.*, 2009; Kotewicz *et al.*, 2007) and whole genome sequence assembly (Wu *et al.*, 2009; Latreille *et al.*, 2007; Zhou *et al.*, 2004; Aston *et al.*, 1999).

Brucellae are a group of Gram-negative, facultative intracellular pathogens that can infect a broad range of mammals, including livestock and humans. The main pathogenic species worldwide are *B. melitensis*, *B. abortus* and *B. suis* that may cause abortion and infertility in their hosts, resulting in vast economic losses. *B. suis* biovars 1, 2 and 3 are the etiological agents involved in swine brucellosis. *B. suis* biovar 2 is also found in hares and biovar 3 in rodents. In Europe, *B. suis* biovar 2 is the most commonly isolated biovar in pigs, wild boars and hares (Grégoire *et al.*, 2012; Godfroid *et al.*, 2011). The prevalence in wild boars appears to be high throughout continental Europe and these animals have been identified as the potential source of transmission of biovar 2 to outdoor or extensively reared pigs (Wu *et al.*, 2012; Godfroid *et al.*, 2011; Muñoz *et al.*, 2010; Galindo *et al.*, 2010). In Portugal and Spain, *B. suis* biovar 2 is the unique biovar isolated from domestic pigs and wild boars (Ferreira *et al.*, 2012; Muñoz *et al.*, 2010).

Molecular characterization by Multi*locus* Variable Number Tandem Repeat Analysis (MLVA) and PCR-restriction fragment length polymorphism (PCR-RFLP) analysis of *omp2a*, *omp2b* and *omp31* identified the existence of two major clonal lineages, one from Central Europe (PCR-RFLP haplotypes 2a, 2b and 2c) and other from Iberian Peninsula (PCR-RFLP haplotypes 2d and 2e), grouped in distinct MLVA clusters (Kreizinger *et al.*, 2014; Muñoz *et al.* 2010; García-Yoldi *et al.*, 2007). Although strains with Central-European haplotype 2c have also been isolated from wild boars in the North-East region of Spain, above Ebro's river, this haplotype had never been reported in domestic pigs in Iberian Peninsula (Muñoz *et al.* 2010).

In this work we compare six *B. suis* biovar 2 strains by WGM to disclose genomic variations among them and to assess the universality of these markers as well as the usefulness of this technology as an epidemiological tool in a large set of field strains.

2. Material and Methods

2.1. *Brucella suis* strains

In this work was used a set of 171 *B. suis* field strains from different hosts and geographic origins, and 22 reference strains representative of the five biovars of *B. suis*, and *B. melitensis* (biovars 1, 2 and 3), *B. abortus* (biovars 1, 2, 3, 4, 5, 6 and 9), *B. ovis*, *B. canis*, *B. ceti*, *B. pinnipedialis* and *B. microti* (**Supplementary Table S3.1**). Field strains outside from Portugal were collected in the framework of *Brucella suis* ring-trial 2006 (COST 845 - Brucellosis in man and animals) and reference strains were obtained from the American Type Culture Collection (ATCC, USA) and National Collection of Type Cultures (NCTC, UK). All strains were maintained in our laboratory at -80 ^{o}C and previously characterized at biovar level, according to standard bacteriological procedures (Alton *et al.*, 1988), as well as at haplotype level by PCR-RFLP analysis of *omp31*, *omp2a* and *omp2b* genes (Vizcaíno *et al.*, 1997; Cloeckaert *et al.*, 1995).

2.2. Optical map construction and sequence-to-map comparison

Genomic DNA was extracted from methanol-treated cells of five *B. suis* biovar 2 field strains, PT09172 (haplotype 2d), PT09143 and Bs143CITA (haplotype 2e), Bs364CITA and Bs396CITA (haplotype 2c), and consensus optical maps were generated at OpGen Technologies, Inc. (Madison, USA) from the restriction patterns obtained at single-molecule level with *BamH* I, using the Argus Whole Genome Mapping System. For comparison, a *BamH* I consensus optical map was also produced for the reference strain *B. suis* biovar 2 strain Thomsen (subsequently referred only as ATCC 23445; haplotype 2a). DNA sequences of chromosome I and II from ATCC 23445 (CP000911, CP000912) and *B. suis* biovar 1 strain 1330 (ATCC 23444; CP002997, CP002998) were downloaded from National Center for Biotechnology Information (NCBI) and *in silico BamH* I restriction maps for each chromosome were created using MapSolver version 3.2 (OpGen Technologies, Inc.). The comparisons between the six optical maps were accomplished by aligning ATCC 23445 and ATCC 23444 *in silico* restriction maps with the optical maps, using the comparative genomics function of MapSolver. Clustering analysis of whole-genome maps (WGMs) was performed with an evaluation license of BioNumerics version 7.1 (kindly provided by Applied Maths, Belgium), using the proportion of matched

fragments as WGMs similarity measure and unweighted pair group method with arithmetic average (UPGMA) as the agglomerative algorithm.

2.3. PCR assessment of chromosomal inversion and indel event

By exploiting the genome sequence data of ATCC 23445, PCR primers were designed at both ends of the inverted region as well as of the 3.5 kb insertion, to assess their presence in all 176 *B. suis* strains. Putative inversion junction regions were identified based on alignment of the restriction pattern and six PCR primers, three flanking each of the putative inversion junction region, were designed and two multiplex PCRs, one for the left (primers F1, R1 and R2) and other for the right (primers F3, R3 and F4) flanking regions, were set up to (see **Figure 3.3a and 3.3b**). Regarding the 3.5 kb indel, two specific primers were designed, IndelF (5´-GTCATCACGCTCCAGGTCTT) and IndelR (5'-CCTGCACACATCAGAACGTC). Specificity of all inversion and insertion targeting primers was confirmed by BLAST analysis against ATCC 23445 and *B. suis* strain 1330 genome sequences. Multiplex-PCR amplifications targeting the inversion and PCR reaction targeting the inversion were performed with MyCycler thermocycler (Bio-Rad, France) in a total volume of 25 µl containing 25 ng of DNA, 10x PCR reaction buffer, 2.5 mM $MgCl_2$, 1 U of Taq DNA polymerase (Promega, USA), 200 mM of each dNTPs and 0.3 mM of each flanking primer. An initial denaturation step at 95 ^{o}C for 5 min was followed by 35 cycles of denaturation at 95 ^{o}C for 30 s, primer annealing at 58 ^{o}C for 30 s and elongation at 72 ^{o}C for 1 min. The final extension step was performed at 72 ^{o}C for 5 min. All PCR products were resolved in a 1.8-2.0% standard agarose gel in 1X Tris-Borate-EDTA buffer, using a 100 bp ladder (Invitrogen, USA) as molecular size marker.

3. Results and Discussion

3.1. Genomic properties and comparison of the six *B. suis* biovar 2 optical maps

Whole-genome mapping (WGM) has a highly discriminant fingerprinting ability (Miller *et al.*, 2013) and, by disclosing the large-scale structure of a genome, can identify genomic changes, such as insertions, deletions, inversions, and translocations events. This technology was applied to assess genomic relatedness in a set of five *B. suis* biovar 2 field strains and compare their genomes with the one of the reference strain for this biovar (ATCC 23445). Optical maps are displayed as a circle showing the position of cleavage sites. Restriction fragments were reassembled into overlapping contiguous regions to create a closed circular map, and the consensus map was built with a minimal coverage of 50X (**Supplementary Figure S3.1**). *BamH* I restriction profiles of chromosome I and chromosome II of

strains ATCC 23445, Bs364CITA, Bs396CITA, PT09172, PT09143 and Bs143CITA, are presented in **Supplementary Table S2** and quantitative analysis was performed by tabulating changes in restriction fragments relative to the reference genome (ATCC 23445). WGM detected 228 to 231 restriction fragments distributed in two chromosomes with fragment size per strain ranging from 1.2 to 85 kb. Strains PT09172 and Bs143CITA lost one restriction site in chromosome I; in chromosome II, PT09172 and Bs396CITA lost two restriction sites and Bs364CITA lost one. These differences were consequence of *BamH* I polymorphisms or failure to digest at the respective positions, which created a net loss of fragments by its fusion. The remaining fragments were comparable in size and position (**Supplementary Tables S3.1A and S3.1B**). All genomes were very similar in size and the estimated chromosome sizes for the six strains ranged from 1,923,017 bp (ATCC 23445) to 1,931,579 bp (Bs364CITA) in chromosome I, and from 1,401,295 bp (ATCC 23445) to 1,404,445 bp (Bs364CITA) in chromosome II (**Table 3.1**).

Table 3.1. Estimated size for each chromosome and respective number of restriction fragments in the optical maps of *B.suis* strains.

***B. suis* strain**	**Clonal lineage[a] / Haplotype[b]**	**Chromosome I**		**Chromosome II**		**Whole-genome**	
		Size (bp)	**Number of fragments**	**Size (bp)**	**Number of fragments**	**Size (bp)**	**Number of fragments**
ATCC 23445 *in silico*		1,923,763	145	1,400,844	98	3,324,607	243
ATCC 23445	CE/ 2a	1,923,017	139	1,401,295	92	3,324,312	232
Bs364CITA	CE/ 2c	1,931,579	139	1,404,445	91	3,336,024	230
Bs396CITA	CE/ 2c	1,923,966	139	1,403,111	90	3,327,077	229
PT09172	IB/ 2d	1,926,303	138	1,402,041	90	3,328,344	228
PT09143	IB/ 2e	1,929,739	139	1,402,229	92	3,331,968	231
Bs143CITA	IB/ 2e	1,929,039	138	1,403,543	92	3,332,582	230

[a] CE, Central-European clonal lineage; IB, Iberian clonal lineage.

[b] Haplotype defined based on the restriction patterns in PCR-RFLP analysis of *omp2a*, *omp2b* and *omp31* genes.

The six closely-related *B. suis* biovar 2 optical maps, and ATCC 23445 and *B. suis* 1330 *in silico* restriction maps, were used to create a map similarity cluster, based on the analysis of 482 fragments classes among strains (**Figure 3.1**). Cluster analysis gathered all biovar 2 strains in a different cluster from the reference strain *B. suis* 1330 (biovar 1) displaying 76.3 % and 79.7 % similarity in chromosome I and II maps, respectively. Considering chromosome I optical maps, strains PT09172, PT09143 and Bs143CITA (haplotypes 2d and 2e) grouped together and in a different cluster from strains from haplotypes 2a, 2b and 2c (97.6 % similarity). Relatively to chromosome II, biovar 2 optical maps were

highly closely-related showing low divergence between strains and haplotypes. Nevertheless, strains PT09143 and Bs143CITA (haplotype 2e) cluster together and the clustering of PT09172 and Bs396CITA demonstrated clear genomic similarities between members of two different haplotypes.

Pairwise alignments were analyzed for differences and compared to ATCC 23445 restriction maps generated *in silico* from the sequence data using MapSolver, assuming the default settings. The comparison of the six chromosome I optical maps revealed the presence of one indel with approximately 3.5 kb in all strains, relatively to ATCC 23445, and the identification of one genetic inversion event in PT09143, PT09172 and Bs143CITA (**Figure 3.2**). This large inversion of approximately 944.2 kb over 1.9 Mbp (49%) on chromosome I, starting at nucleotide #371383, and encompassing 961 genes in ATCC 23445, showed similar restriction profiles in the three strains from haplotypes 2d and 2e. Until now and to the best of our knowledge, only one large chromosomal inversion of 640 kb was described in *B. abortus* within the members of genus *Brucella* (Halling *et al.*, 2005).

The presence of the large genetic inversion in *B. suis* was assessed by multiplex-PCR targeting the left- and right-junction sites (**Figure 3.3**) on a total of 176 *B. suis* strains, including all haplotypes, and on 15 reference strains from *B. melitensis*, *B. abortus, B. ovis*, *B. canis*, *B. ceti*, *B. pinnipedialis* and *B. microti*. The inversion was confirmed in all strains from Iberian clonal lineage (haplotypes 2d and 2e) but not in Central-European strains (haplotypes 2a, 2b and 2c) or any other *B. suis* biovars or *Brucella* species (**Figure 3.3c**; **supplementary Table S3.2**). Considering these results, it seems evident that the inversion occurred in the ancestor of the Iberian clonal lineage of *B.suis* biovar 2.

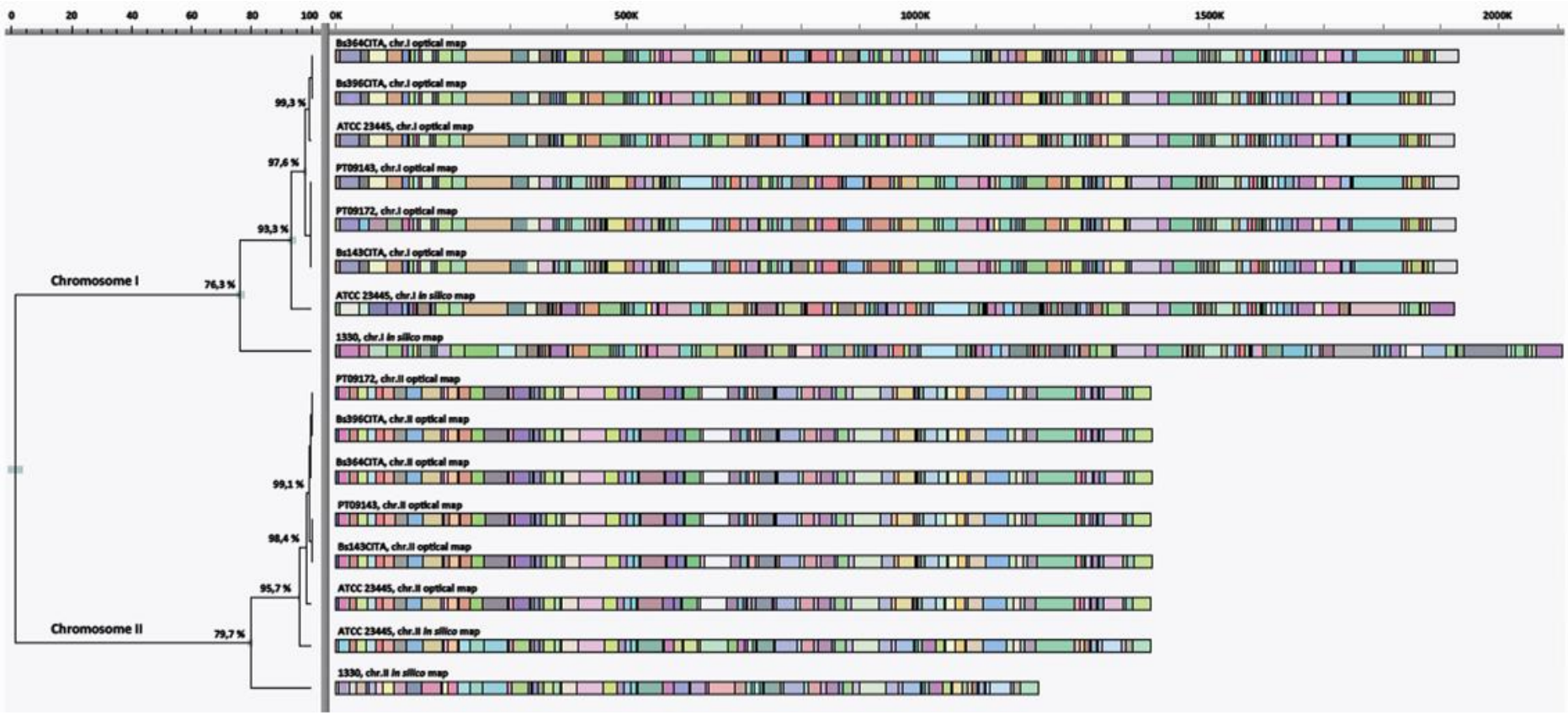


Figure 3.1. Map similarity cluster of chromosomes I and II of *B. suis* strains using unweighted pair group method with arithmetic average (UPGMA). The cluster tree is drawn with the lengths of branches indicating the relative similarity between two nodes. A similarity ruler is shown at the top of the tree for comparing the differences between maps. Whole genome *BamH* I *in silico* and optical maps of *B. suis* strains are represented as linear maps displaying the fragments in randomly chosen colors, with matching fragments sharing the same color. The clustering displayed a total of 482 fragments classes among strains. *B. suis* biovar 1 strain 1330 was used as an outgroup (*in silico* maps).

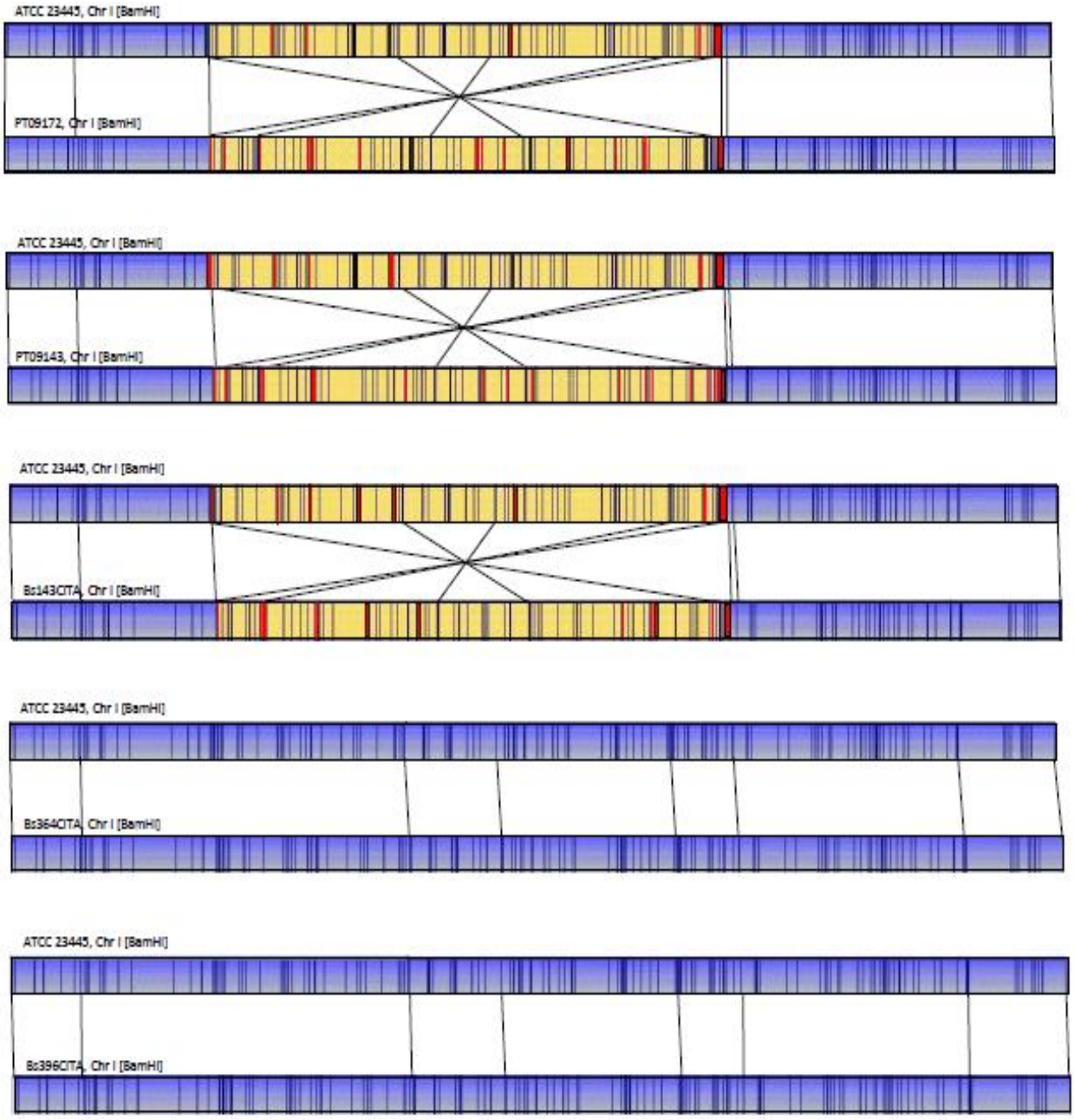
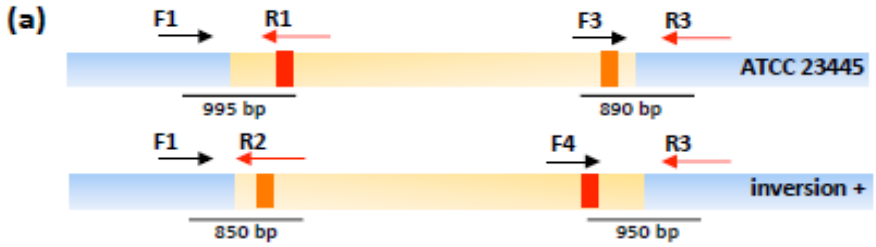


Figure 3.2. Pairwise alignment of chromosome I optical maps for *B. suis* biovar 2 Iberian strains and *in silico* reference map. Lines connecting two chromosomal maps indicate discontinuity in the alignment of fragments. The chromosomal inversion relative to strain ATCC 23445 is highlighted in yellow and is indicated by crossed alignment lines between paired maps; unaligned restriction fragments, representing differences between two chromosomes, are showed in red; blue indicates aligned restriction fragments.



(b)

Primer	Forward (F) or Reverse (R)	Sequence (5´– 3´)	PCR product size (bp)	
			Iberian strains	Others
Multiplex-PCR left region				
F1	F	TCTTCTGCCCGTCAATCTGT	850	995
R1	R	CGCCAAACCAGATGAGATCG		
R2	R	GCCGAGACAGAGGATAACGA		
Multiplex-PCR right region				
F3	F	GAGGATAACGACGGTGGGAT		
R3	R	AACAAGGTTCATCGCGCTTT	950	890
F4	F	AGGCCAGAAAGAACACCAGA		

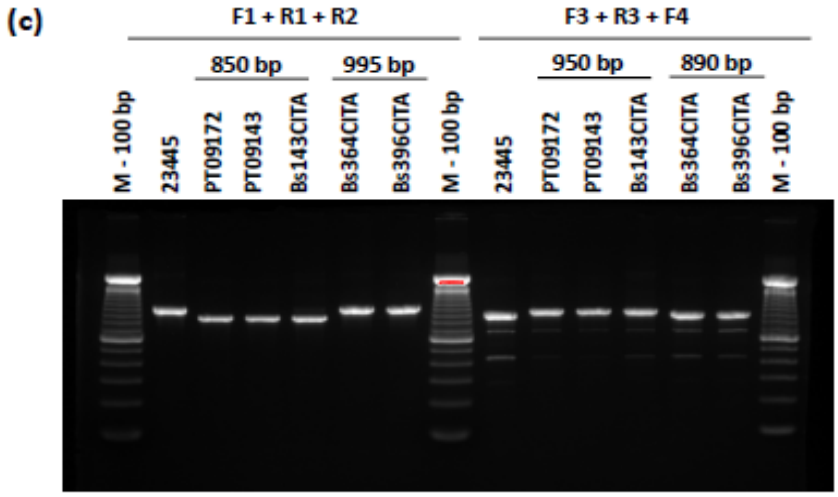


Figure 3.3. Assessment of the large genomic inversion by left- and right-junction site-specific PCR. **(a)** Schematic representation of the relative position of primers and expected amplicon sizes for strain ATCC 23445 (without inversion) and an inversion positive strain. **(b)** Primers used in multiplex-PCR targeting the left- and right regions of the chromosomal inversion and product size obtained with strains from the Iberian clonal lineage (inversion positive) and other strains. **(c)** Multiplex-PCR analysis of the six strains used for optical maps construction: ATCC 23445 (haplotype 2a), PT09172 (haplotype 2d), PT09143 and Bs143CITA (haplotype 2e), Bs364CITA and Bs396CITA (haplotype 2c). M: 100 bp DNA size marker (Invitrogen).

Similarly, the presence/absence of the 3.5 kb indel in chromosome I was searched by PCR and it was confirmed that this indel is only present in the reference strain ATCC 23445 (**supplementary Table S3.2**). BLAST searches against *Brucella* published genomes using the nucleotide sequence of the indel region in the genome of *B. suis* strain 1330 (ATCC 23444) as a query, identified a putative esterase (BR0127), a putative transmembrane protein (BR0128), an acetylransferase (BR0129), a S-(hydroxymethyl)-glutathione dehydrogenase (pseudogene BR0130) and a transcriptional regulator

(BR0131). Homologous genes are present in all *Brucella* species except *B. suis* strain ATCC 23445, confirming the absence of this indel region in the genome sequence of this last strain.

Regarding chromosome II optical maps, no relevant insert/deletion or other genomic rearrangements events were found but few missing cuts occurred leading to the fusion of fragments (**Supplementary Table S3.1A and S3.1B**).

The comparison of optical maps of new isolates with *in silico* optical maps of characterized strains can identify potential functional, or phenotypic differences between strains, generating empirical markers to distinguish between identical isolates, providing a powerful tool for epidemiological studies (Kotewicz *et al.*, 2007). In this work it is shown that optical maps are appropriate to distinguish closely-related *B. suis* strains based on the large number of restriction sites analyzed, allowing the discrimination between different strains of the same biovar and haplotype, making this method a very accurate and useful tool to investigate transmission events and outbreaks involving pathogens of this species.

3.2. Comparison of *BamH* I *in silico* and optical maps of ATCC 23445

Optical maps of ATCC 23445 chromosome I and II were compared to the restriction maps generated *in silico* from the DNA sequences CP000911 and CP000912. WGM detected 139 out of 145 and 92 out of 98 of expected fragments (95.1%) in chromosome I and II, respectively. With the exception of the six fragments that were not detected in each chromosome (with less than 600 bp in size), the optical maps recapitulate the sequence-based map. In comparison to the reference DNA sequences of ATCC 23445, the optical map showed an estimated loss of 746 bp (0.04%) and a gain of 451 bp (0.03%) in chromosome I and II, respectively (**Supplementary Table S3.1A and S3.1B**). The pairwise comparisons of the restriction maps exposed two regions of misalignment: one region in chromosome I (8,355 bp) and one in chromosome II (46624 bp) (**Figure 3.4**). The reference strain ATCC 23445 used to produce the optical map is the same strain that was used to generate the available whole-genome sequence, and from which the *in silico* maps derived. However, frequent subculturing can lead to selection of clone variants that accumulate point mutations occurring over time and this may explain some of the observed differences between map profiles. On the other hand, the expected resolution of optical mapping is limited to fragments over *ca.* 600 bp and so the restriction sites delimiting fragments below 750 bp (6 in each chromosome) could not be included in the optical map, contributing for the differences observed between *in silico* and optical map of reference strain. Additionally, sequencing and assembly errors at DNA sequences cannot be ignored as a potential source for further differences. However, previously published works have shown that the resolution and detail that optical maps provide underscore their usefulness for molecular epidemiological studies

and its utility to complement genome sequencing projects (Sabirova *et al.* 2014; Schwan *et al.*, 2010; Shukla *et al.*, 2009; Wu *et al.*, 2009; Kotewicz *et al.*, 2007; Zhou *et al.*, 2004).

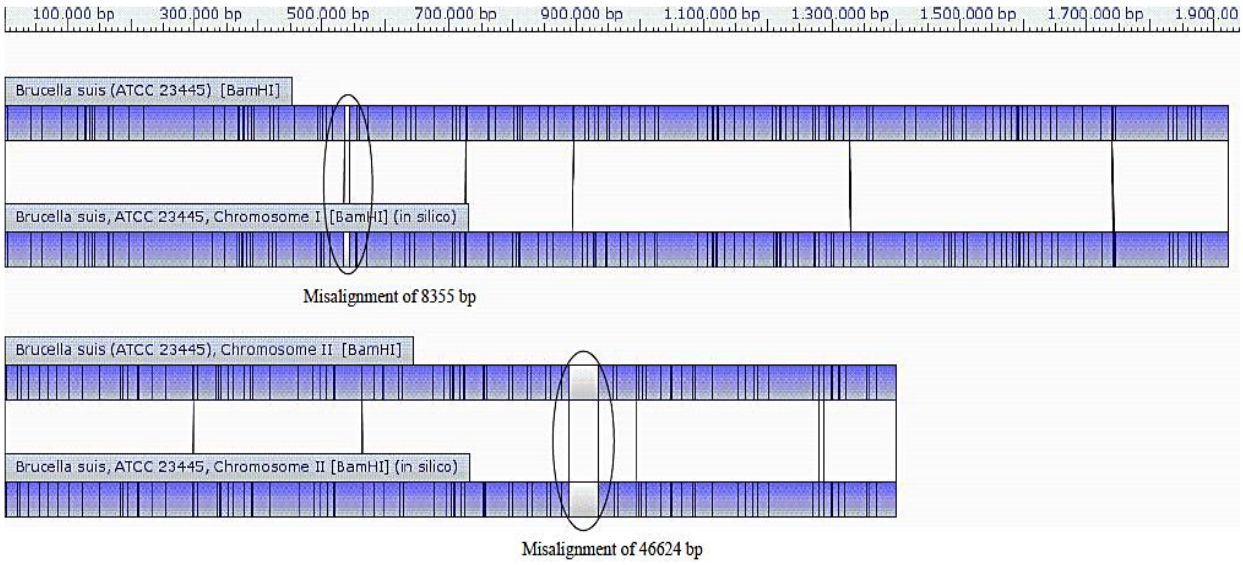


Figure 3.4. Comparison of the *BamH I* optical and *in silico* maps of *B. suis* ATCC 23445. A reference bp scale is present on top of the maps. There are 145 *BamH* I restriction fragments in chromosome I *in silico* map and 139 fragments in the optical map; in chromosome II, the optical map contains 92 *BamH* I fragments and the *in silico* map contains 98 fragments. The black lines linking maps are the alignment connectors. Similarities between maps are displayed in blue and differences (unaligned restriction fragments) in white. The regions of misalignment in chromosome I and II are indicated.

4. Conclusion

Beyond the higher strain typing resolution observed in this work with WGM, the comparative genomic analysis between optical and *in silico* maps allowed the identification of a large genomic inversion event in *Brucella suis* strains with haplotypes 2d and 2e, so far only isolated in Portugal and Spain and belonging to the Iberian clonal lineage. Data obtained with WGM support thus the existence of the Central-European clonal lineage (haplotypes 2a, 2b and 2c) and the Iberian clonal lineage (haplotypes 2d and 2e) already identified by MLVA and PCR-RFLP of *omp* genes. Molecular identification of both clonal lineages can be easily achieved using the multiplex-PCR procedures developed in this work and targeting the left- or the right-junction sites of the chromosomal inversion.

5. Supplementary material

Supplementary Figure S3.1. Whole-genome *BamH* I optical maps of chromosome I and chromosome II of the six *B. suis* biovar 2 strains: ATCC 23445, Bs364CITA, Bs396CITA, PT09172, PT09143 and Bs143CITA.

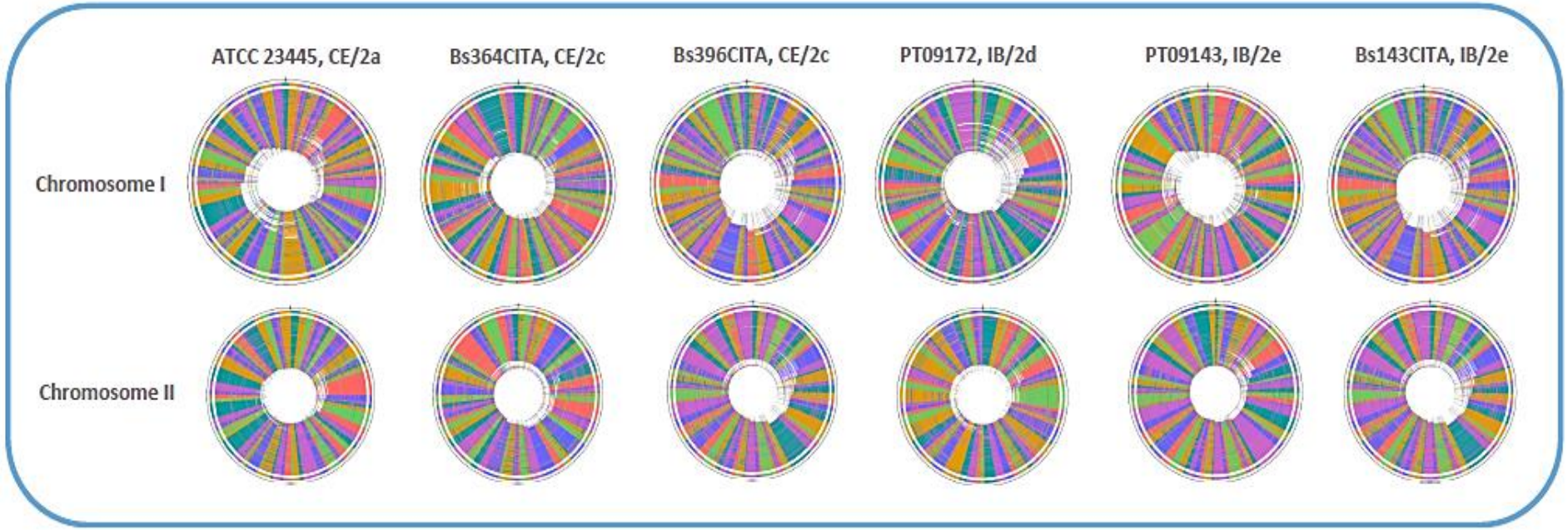
Supplementary Table S3.1. **(A)** Chromosome I restriction fragments for each *B. suis* biovar 2 *BamH* I *in silico* and optical maps; **(B)** Chromosome II restriction fragments for each *B. suis* biovar 2 *BamH* I *in silico* and optical maps.

Supplementary Table S3.2. *Brucella* strains used to search for the inversion in chromosome I and the large indel event in the reference strain *B. suis* ATCC 23445.

6. References

1. **Anantharaman, T. & Mishra, B. (2013**). Genomics via Optical Mapping (I): 0-1 Laws for Mapping with Single Molecules. BioRxiv, doi: http://dx.doi.org/10.1101/000844.
2. **Alton, G.C., Jones, L.M., Angus, R.D. & Verger, J.M. (1988**). Techniques for the Brucellosis Laboratory. Institut National de la Recherche Agronomique (INRA), Paris, France.
3. **Aston, C., Mishra, B. & Schwartz, D.C. (1999**). Optical mapping and its potential for large-scale sequencing projects. Trends in Biotechnology, 17: 297-302.
4. **Bosch, T., Verkade, E., van Luit, M., Pot, B., Vauterin, P., Burggrave, R., Savelkoul, P., Kluytmans, J. & Schouls, L. (2013)**. High resolution typing by whole genome mapping enables discrimination of LA-MRSA (CC398) strains and identification of transmission events. PLoS ONE, 8(6): e66493. doi:10.1371/journal.pone.0066493. doi:10.1371/journal.pone.0066493.
5. **Cloeckaert, A., Verger, J.M., Grayon, M. & Grépinet, O. (1995)**. Restriction site polymorphism of the genes encoding the major 25 and 36 kDa outer-membrane proteins of *Brucella*. Microbiology, 141: 2111–2121.
6. **Ferreira, A.C., Almendra, C., Cardoso, R., Silva-Pereira, M., Beja-Pereira, A., Luikart, G. & De Sá, M.I.C. (2012)**. Development and evaluation of a selective medium for *Brucella suis*. Research in Veterinary Science, 93: 565–567.
7. **Galindo, R.C., Muñoz, P.M, De Miguel, M.J., Marin, C.M., Labairu, J., Revilla, M., Blasco, J.M., Gortazar, C. & de la Fuente, J. (2010)**. Gene expression changes in spleens of the wildlife reservoir species, Eurasian wild boar (*Sus scrofa*), naturally infected with *Brucella suis* biovar 2. Journal of Genetics and Genomics, 37: 725-736.
8. **García-Yoldi, D., Le Flèche, P., de Miguel, M.J., Muñoz, P.M., Blasco, J.M., Cvetnic, Z., Marín, C.M., Vergnaud, G. & López-Goñi, I. (2007)**. Comparison of multiple-*locus* variable-number tandem-repeat analysis with other PCR-based methods for typing *Brucella suis* isolates. Journal of Clinical Microbiology, 45: 4070–4072.
9. **Godfroid, J., Scholz, H.C., Barbier, T., Nicolas, C., Wattiau, P., Fretin, D., Whatmore, A.M., Cloeckaert, A., Blasco J.M., Moriyon, I., Saegerman, C., Mumak, J.B., Al Dahouk, S., Neubauer, H. & Letesson, J.-J. (2011)**. Brucellosis at the animal/ecosystem/human interface at the beginning of the 21st century. Preventive Veterinary Medicine, 102(2): 118-31. doi: 10.1016/j.prevetmed.2011.04.007.
10. **Grégoire, F., Mousset, B., Hanrez, D., Michaux, C., Walravens, K. & Linden, A. (2012).** A serological and bacteriological survey of brucellosis in wild boar (*Sus scrofa*) in Belgium. BMC Veterinary Research, 8: 80.

11. **Halling, S.M., Peterson-Burch, B.D., Bricker, B.J., Zuerner, R.L., Qing, Z., Li, L.L., Kapur, V., Alt, D.P. & Olsen, S.C. (2005)**. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella* melitensis and *Brucella* suis. Journal of Bacteriology, 187 (8): 2715-2726.
12. **Kreizinger, Z., Foster, J.T., Rónai, Z., Sulyok, K.M., Wehmann, E., Jánosi, S. & Gyuranecz, M. (2014)**. Genetic relatedness of *Brucella suis* biovar 2 isolates from hares, wild boars and domestic pigs. Veterinary Microbiology, 172: 492-8. doi: 10.1016/j.vetmic.2014.05.031.
13. **Kotewicz, M.L., Jackson, S.A., LeClerc, J.E. & Cebula, T.A. (2007)**. Optical maps distinguish individual strains of *Escherichia coli* O157: H7. Microbiology, 153: 1720-1733.
14. **Latreille, P., Norton, S., Goldman, B., Henkhaus, J., Miller, N., Barbazuk, B., Bode, H., Darby, C., Du, Z., Forst, S., Gaudriault, S., Goodner, B., Goodrich-Blair, H. & Slater, S. (2007)**. Optical mapping as a routine tool for bacterial genome sequence finishing. BMC Genomics, 8: 321.
15. **Miller, J.M. (2013**). Whole-Genome Mapping: a New Paradigm in Strain-Typing Technology. Journal of Clinical Microbiology, 51: 1066-1070.
16. **Muñoz, P.M., Boadella, M., Arnal, M., de Miguel, M.J., Revilla, M., Martínez, D., Vicente, J., Acevedo, P., Oleaga, A., Ruiz-Fons, F., Marín, C.M., Prieto, J.M., de la Fuente, J., Barral, M., Barberán, M., de Luco, D.F., Blasco, J.M. & Gortázar, C. (2010)**. Spatial distribution and risk factors of brucellosis in Iberian wild ungulates. BMC Infectious Diseases, 10: 46.
17. **Sabirova, J.S., Xavier, B.B., Ieven, M., Goossens, H. & Malhotra-Kumar, S. (2014).** Whole genome mapping as a fast-track tool to assess genomic stability of sequenced Staphylococcus aureus strains. BMC Research Notes, 7: 704.
18. **Schwan, W.R., Briska, A., Stahl, B., Wagner, T.K., Zentz, E., Henkhaus, J., Lovrich, S.D., Agger, W.A., Callister, S.M., DuChateau, B. & Dykes, C.W. (2010)**. Use of optical mapping to sort uropathogenic *Escherichia coli* strains into distinct subgroups. Microbiology, 156: 2124-2135.
19. **Shukla, S.K., Kislow, J., Briska, A., Henkhaus, J. & Dykes, C. (2009)**. Optical mapping reveals a large genetic inversion between two methicillinresistant Staphylococcus aureus strains. Journal of Bacteriology, 191: 5717-5723.
20. **Vizcaíno, N., Verger, J.M., Grayon, M., Zygmunt, M.S. & Cloeckaert, A. (1997)**. DNA polymorphism at the omp-31 *locus* of *Brucella* spp.: evidence for a large deletion in *Brucella abortus*, and other species-specific markers. Microbiology, 143: 2913–2921.
21. **Wu, C.W., Schramm, T.M., Zhou, S., Schwartz, D.C. & Talaat, A.M. (2009).** Optical mapping of the *Mycobacterium avium* subspecies *paratuberculosis* genome. BMC Genomics, 10: 25. doi: 10.1186/1471-2164-10-25.
22. **Wu, N., Abril, C., Thomann, A., Grosclaude, E., Doherr, M.G., Boujon, P. & Ryser-Degiorgis, M.P. (2012)**. Risk factors for contacts between wild boar and outdoor pigs in Switzerland and investigations on potential *Brucella suis* spill-over. BMC Veterinary Research, 8: 116. doi: 10.1186/1471-2164-10-25.
23. **Zhou, S., Herscheleb, J. & Schwartz, D.C. (2007)**. A Single Molecule System for Whole Genome Analysis. New high throughput technologies for DNA sequencing and genomics 2. Elsevier. pp. 269–304.
24. **Zhou, S., Kile, A., Bechner, M., Place, M., Kvikstad, E., Deng, W., Wei, J., Severin, J., Runnheim, R., Churas, C., Forrest, D., Dimalanta, E.T., Lamers, C., Burland, V., Blattner, F.R. & Schwartz, D.C. (2004)**. Single-molecule approach to bacterial genomic comparisons via optical mapping. Journal of Bacteriology, 186(22): 7773-82.



Supplementary Figure S3.1. Whole-genome *BamH I* optical maps of chromosome I and chromosome II of the six *B. suis* biovar 2 strains: ATCC 23445, Bs364CITA, Bs396CITA, PT09172, PT09143 and Bs143CITA.

The clonal lineage (CE, Central-European; IB, Iberian) and the haplotype (2a, 2c, 2d and 2e) are indicated for each strain. Chromosome I and chromosome II *BamH* I optical maps of the six *B. suis* biovar 2 strains are displayed as a circle showing the position of cleavage sites. Each series of concentric multicolor annuli represents optical maps constructed from individual molecules. The consensus map (external circle) was built from the underlying maps within a minimal coverage of 50X. Matching restriction fragments are represented by a common color; the color coding is random to enhance contrast. Missing cuts are displayed as fragments that span across one or more cut sites in the consensus map.

Supplementary Table S3.1. Chromosome I (A) and chromosome II (B) restriction fragments for each *B. suis* biovar 2 *BamH* I *in silico* and optical maps.

(A) Chromosome I restriction fragments

Fragment order number in each profile	ATCC 23445 *in silico* map	ATCC 23445 optical map	Bs364CITA optical map	Bs396CITA optical map	PT09172 optical map	PT09143 optical map	Bs143CITA optical map
#1	6,188	6,668	6,720	6,716	6,793	6,759	6,775
#2	36,822	36,745	37,450	37,821	36,906	36,951	36,670
#3	15,396	15,513	15,690	15,515	15,695	15,589	15,700
#4	30,947	30,501	30,780	30,920	30,193	30,420	30,101
#5	26,589	26,156	26,293	25,980	26,261	26,517	26,329
#6	10,391	10,277	10,263	10,299	11,110	10,534	10,294
#7	750	1,598	1,651	1,487	7,889	1,575	1,597
#8	7,327	7,538	7,377	7,475	8,056	7,395	7,239
#9	4,502	4,617	8,062	8,039	4,015	8,215	8,067
#10	3,743	3,951	3,828	3,971	19,575	4,115	3,976
#11	19,805	19,450	19,600	19,467	4,590	19,620	19,608
#12	4,091	4,352	4,538	4,324	5,185	4,424	4,398
#13	4,919	5,144	5,092	5,152	23,574	5,031	5,125
#14	23,698	23,549	23,457	23,488	23,802	23,515	23,502
#15	23,691	23,800	23,823	23,735	77,717	24,068	23,927
#16	77,587	77,971	77,603	77,757	30,624	78,114	77,268
#17	30,804	30,693	30,674	30,760	18,959	30,859	30,621
#18	19,210	19,103	19,207	18,933	24,141	18,972	19,097
#19	20,738	20,515	20,345	20,323	4,188	24,158	24,258
#20	2,498	2,831	2,846	2,761	6,650	4,165	4,036
#21	4,363	4,534	4,582	4,640	11,089	6,495	6,598
#22	1,663	2,080	2,096	1,972	5,464	11,044	11,134
#23	5,351	5,319	5,590	5,511	3,069	5,482	5,496
#24	6,173	6,353	6,281	6,324	24,127	3,155	3,151
#25	2,913	3,220	3,316	3,222	7,397	24,252	24,049
#26	25,989	25,322	25,315	25,365	10,709	7,427	7,499
#27	4,480	4,976	4,972	4,910	8,461	10,702	10,626
#28	6,980	7,381	7,419	7,366	2,592	8,609	8,669
#29	26,768	26,225	26,300	26,151	4,556	2,510	2,666
#30	36,041	35,439	35,343	35,514	1,818	4,195	4,259
#31	5,059	5,432	5,487	5,316	3,808	1,704	1,811
#32	4,248	4,280	4,352	4,295	29,808	3,951	3,875
#33	4,255	4,525	4,472	4,412	14,782	29,822	29,783
#34	9,296	9,251	9,368	9,456	16,023	14,758	14,945
#35	21,048	20,501	20,543	20,358	14,351	16,010	16,012
#36	8,248	8,355	8,184	8,146	12,092	14,302	14,418
#37	9,238	9,300	9,253	9,107	2,443	12,038	12,099
#38	3,020	3,591	3,562	3,536	4,767	2,520	2,519
#39	15,402	15,466	15,408	15,393	2,158	4,479	4,622

Fragment order number in each profile	ATCC 23445 *in silico* map	ATCC 23445 optical map	Bs364CITA optical map	Bs396CITA optical map	PT09172 optical map	PT09143 optical map	Bs143CITA optical map
#40	38,636	37,833	37,813	37,519	7,656	2,115	2,042
#41	22,014	22,040	22,119	22,128	16,143	7,708	7,855
#42	7,132	7,371	7,427	7,387	60,736	16,115	16,160
#43	6,745	6,838	6,892	6,885	5,963	60,371	60,100
#44	30,304	29,900	29,739	29,681	15,093	6,096	5,975
#45	29,701	29,228	28,974	28,938	7,684	15,023	15,106
#46	4,285	4,517	4,644	4,471	18,401	7,651	7,802
#47	6,373	6,385	6,584	6,420	11,126	18,603	18,501
#48	7,921	8,232	8,214	8,069	19,933	10,986	11,285
#49	1,132	2,121	2,190	1,978	2,586	19,711	20,151
#50	142	33,679	33,593	33,405	14,295	2,599	2,700
#51	34,701	2,781	2,812	2,617	5,389	14,261	14,131
#52	2,323	8,373	8,543	8,335	9,711	5,405	5,392
#53	8,070	30,767	30,801	30,430	6,964	9,666	9,759
#54	31,021	5,347	5,333	5,187	16,324	6,901	7,028
#55	4,719	3,145	3,032	3,059	1,638	16,209	16,200
#56	2,713	2,642	2,791	2,758	28,054	1,663	1,794
#57	2,348	28,405	28,076	27,839	10,774	27,953	27,943
#58	29,031	13,282	13,340	13,113	13,310	10,853	10,824
#59	13,396	10,666	10,881	10,689	28,227	13,272	13,408
#60	10,634	27,991	27,901	27,738	2,816	28,182	28,149
#61	28,747	1,626	1,856	1,636	3,259	2,823	2,753
#62	722	16,250	16,158	16,096	5,183	3,308	3,254
#63	16,544	6,956	7,152	6,906	30,604	5,026	5,255
#64	7,026	9,656	9,609	9,806	8,347	30,519	30,449
#65	9,576	5,444	5,515	5,326	2,726	8,421	8,359
#66	681	14,131	14,453	14,160	33,660	2,732	2,837
#67	4,815	2,639	2,591	2,631	1,919	33,671	33,594
#68	14,440	19,919	19,967	19,840	8,181	2,030	2,059
#69	2,142	10,936	11,047	10,938	6,423	8,171	8,200
#70	20,414	18,411	18,439	18,430	4,534	6,458	6,503
#71	11,188	7,736	7,827	7,774	28,850	4,423	4,711
#72	19,047	14,969	15,067	14,839	29,416	29,059	29,149
#73	7,660	5,964	6,024	5,954	6,660	29,816	29,864
#74	15,379	60,321	60,255	60,343	7,256	6,820	6,865
#75	5,558	16,064	16,176	15,993	21,765	7,307	7,334
#76	61,207	7,683	7,688	7,624	37,525	21,990	21,835
#77	16,195	2,201	2,384	2,248	15,254	37,776	37,636
#78	7,553	4,464	4,666	4,595	3,566	15,416	15,426
#79	1,825	2,484	2,474	2,550	9,208	3,480	3,491
#80	4,335	11,965	12,274	11,937	8,214	9,171	9,350
#81	2,296	14,288	14,332	14,203	20,300	8,254	8,286
#82	12,457	16,038	16,067	16,077	9,200	20,505	20,375
#83	14,578	14,762	15,038	14,819	4,368	9,371	9,509
#84	16,311	29,460	29,740	29,687	4,389	4,303	4,383

Fragment order number in each profile	ATCC 23445 *in silico* map	ATCC 23445 optical map	Bs364CITA optical map	Bs396CITA optical map	PT09172 optical map	PT09143 optical map	Bs143CITA optical map
#85	14,947	3,796	3,955	3,909	5,394	4,376	4,312
#86	30,492	1,799	1,826	1,669	35,136	5,474	5,536
#87	3,499	4,426	4,459	4,342	26,202	35,325	35,611
#88	149	2,572	2,484	2,604	7,198	26,163	26,204
#89	744	8,620	8,613	8,422	4,854	7,312	7,370
#90	3,890	10,794	10,701	10,692	25,112	4,862	5,062
#91	2,544	7,341	7,520	7,490	3,186	25,265	25,193
#92	8,571	23,834	24,006	24,135	6,324	3,278	3,392
#93	10,879	3,106	3,234	3,126	5,264	6,219	6,309
#94	7,581	5,443	5,404	5,335	1,921	5,498	5,345
#95	24,950	11,007	11,030	10,862	4,404	1,832	2,045
#96	2,812	6,571	6,444	6,481	2,671	4,255	4,381
#97	5,138	3,760	3,888	3,864	11,528	2,719	2,681
#98	11,188	15,252	15,263	15,411	11,213	11,641	11,983
#99	6,554	11,233	11,304	11,155	25,774	11,281	11,329
#100	3,725	26,013	26,104	25,975	3,632	26,030	25,912
#101	15,698	3,695	3,835	3,785	5,865	3,766	3,922
#102	11,157	5,857	5,887	5,895	48,691	5,947	6,081
#103	210	48,966	48,418	48,758	19,821	48,726	48,923
#104	26,167	19,869	19,972	19,839	41,310	19,927	19,883
#105	3,286	41,347	41,371	41,793	7,003	41,559	41,706
#106	5,466	7,010	7,159	6,934	4,520	7,075	7,103
#107	49,516	4,521	4,455	4,409	6,874	4,500	4,443
#108	19,884	6,890	6,802	6,841	12,025	6,866	6,918
#109	42,337	12,053	12,062	12,062	6,508	12,097	12,019
#110	6,769	6,491	6,520	6,428	30,343	6,484	6,590
#111	4,163	30,304	30,373	30,285	10,243	30,346	30,355
#112	6,640	10,240	10,316	10,276	14,083	10,386	10,216
#113	12,110	13,971	13,884	13,944	6,561	13,988	13,953
#114	6,200	6,638	6,588	6,625	9,477	6,583	6,693
#115	31,207	9,498	9,640	9,610	6,865	9,429	9,299
#116	10,176	6,951	7,277	6,935	2,869	7,188	8,535
#117	14,109	2,693	2,734	2,790	2,868	2,605	2,971
#118	6,538	3,041	2,981	3,043	4,708	3,040	5,797
#119	9,761	4,662	4,653	4,630	8,497	4,567	8,432
#120	7,051	8,488	8,587	8,507	9,908	8,474	9,952
#121	2,631	9,950	10,050	9,931	10,348	10,000	10,308
#122	2,898	10,200	10,239	10,307	15,501	10,266	15,410
#123	4,420	15,479	15,584	15,489	9,535	15,635	9,513
#124	8,462	9,580	9,521	9,398	27,214	9,513	26,970
#125	10,205	27,004	27,000	27,080	14,688	27,017	14,560
#126	10,378	14,665	14,593	14,529	27,274	14,691	27,184
#127	15,380	27,130	27,164	27,290	1,903	27,435	1,969
#128	9,474	1,986	1,967	1,875	16,206	1,937	15,864
#129	27,636	15,976	15,990	15,858	2,037	16,077	2,046

Fragment order number in each profile	ATCC 23445 *in silico* map	ATCC 23445 optical map	Bs364CITA optical map	Bs396CITA optical map	PT09172 optical map	PT09143 optical map	Bs143CITA optical map
#130	14,528	2,196	2,285	2,134	3,005	2,150	3,162
#131	27,787	3,050	3,249	3,104	85,094	3,084	84,104
#132	1,063	83,850	83,493	83,963	7,013	84,538	6,955
#133	15,938	6,929	7,045	6,898	6,506	6,940	6,507
#134	1,423	6,429	6,498	6,498	18,684	6,490	18,319
#135	2,124	18,290	18,459	18,509	4,344	18,678	4,371
#136	84,786	4,368	4,382	4,282	6,879	4,381	6,827
#137	6,627	6,793	6,844	6,882	7,790	6,895	7,926
#138	6,257	7,889	7,836	7,967	44,363	7,962	43,987
#139	18,560	44,069	43,991	44,136		44,323	
#140	3,946						
#141	162						
#142	6,204						
#143	455						
#144	7,144						
#145	44,899						
Sum of fragments size	1,923,763	1,923,017	1,931,579	1,923,966	1,926,303	1,929,739	1,929,039
Min fragment size	142	1,598	1,651	1,487	1,638	1,575	1,597
Max fragment size	84,786	83,850	83,493	83,963	85,094	84,538	84,104
Average fragment size	13,267	13,835	13,896	13,841	13,959	13,883	13,979
Number of fragments	145	139	139	139	138	139	138

(B) Chromosome II restriction fragments

Fragment order number in each profile	ATCC 23445 *in silico* map	ATCC 23445 optical map	Bs364CITA optical map	Bs396CITA optical map	PT09172 optical map	PT09143 optical map	Bs143CITA optical map
#1	4,074	4,333	4,458	4,383	4,395	4,428	4,433
#2	17,599	17,408	17,303	17,506	17,583	17,612	17,406
#3	3,740	4,290	4,394	4,204	4,275	4,338	4,335
#4	12,842	12,901	12,880	12,917	12,805	12,955	12,876
#5	17,223	16,806	16,770	16,880	16,740	16,808	16,847
#6	14,745	14,517	14,548	14,558	14,540	14,631	14,572
#7	13,619	13,471	13,546	13,461	13,450	13,597	13,569
#8	18,482	18,553	18,486	18,485	18,343	18,465	18,528
#9	21,350	21,379	21,384	21,372	21,513	21,601	21,564
#10	26,465	26,166	26,249	26,338	26,199	26,535	26,387
#11	33,190	32,658	32,526	32,718	32,791	32,747	32,657
#12	2,842	3,229	3,374	3,297	3,239	3,282	3,361
#13	8,445	8,332	8,475	8,352	8,415	9,322	9,420
#14	15,839	15,735	15,943	15,941	15,781	15,908	15,823
#15	300	2,173	2,276	2,121	2,166	2,189	2,232
#16	1,162	20,092	20,285	20,285	20,340	20,317	20,231
#17	20,333	21,558	21,847	21,851	21,672	21,854	21,962
#18	21,788	41,728	41,606	41,830	41,696	41,920	41,873
#19	42,392	1,615	1,770	1,356	1,477	1,685	1,581
#20	618	7,503	7,528	7,499	7,451	7,551	7,710
#21	7,095	30,038	30,280	30,325	30,518	30,347	30,512
#22	31,236	3,662	3,734	3,687	3,639	3,618	3,664
#23	3,324	3,734	3,979	3,852	3,762	3,776	3,725
#24	3,324	9,681	9,860	9,844	9,764	9,642	9,713
#25	9,688	8,150	8,191	8,151	8,136	8,182	8,217
#26	8,302	17,236	17,339	17,396	17,297	17,319	17,309
#27	17,542	10,369	10,527	10,535	10,307	10,312	10,462
#28	10,137	2,835	2,922	2,868	2,795	2,828	2,925
#29	2,477	27,341	27,457	27,730	27,800	27,622	27,701
#30	27,920	43,860	44,227	44,826	44,027	44,062	44,473
#31	44,635	23,585	23,712	23,762	23,618	23,562	23,666
#32	23,877	11,107	11,790	11,787	11,791	11,592	11,785
#33	11,079	11,604	11,672	11,769	11,595	11,416	11,582
#34	11,483	9,328	9,436	9,579	9,267	9,302	9,356
#35	9,071	2,080	2,152	2,230	2,105	2,142	2,212
#36	1,374	42,595	42,252	43,206	42,990	42,149	42,244
#37	43,502	18,996	1,487	19,154	18,954	1,336	1,584
#38	474	13,162	18,721	13,307	13,201	18,607	18,473
#39	18,742	1,967	13,164	1,995	1,878	13,110	13,288
#40	13,155	24,835	2,050	25,164	25,084	1,939	2,107
#41	1,218	4,936	25,101	4,974	4,904	24,906	24,943
#42	25,767	48,417	5,030	49,001	47,886	4,834	5,049
#43	4,364	17,474	48,295	17,745	17,557	47,190	47,739
#44	48,947	10,345	17,721	10,347	10,415	17,446	17,792
#45	17,741	4,107	10,384	3,892	3,913	10,284	10,376

Fragment order number in each profile	ATCC 23445 *in silico* map	ATCC 23445 optical map	Bs364CITA optical map	Bs396CITA optical map	PT09172 optical map	PT09143 optical map	Bs143CITA optical map
#46	10,449	1,881	3,954	10,093	10,083	3,886	4,035
#47	3,913	8,797	10,111	6,567	6,501	10,001	9,887
#48	1,340	6,027	6,709	28,457	28,556	6,524	6,601
#49	8,735	1,258	28,183	2,209	2,094	27,997	28,178
#50	5,689	27,891	2,156	3,796	3,753	2,075	2,175
#51	769	2,129	3,791	34,766	34,935	3,678	3,749
#52	28,719	3,480	34,938	7,138	7,319	34,795	34,736
#53	1,011	35,174	7,424	21,473	21,985	7,145	7,294
#54	2,979	7,254	21,717	8,257	8,382	21,673	21,562
#55	36,024	21,664	8,436	21,578	21,869	8,390	8,416
#56	7,124	8,351	21,721	6,445	6,561	21,689	21,799
#57	21,854	21,879	6,547	17,292	17,517	6,504	6,519
#58	7,973	6,543	17,292	11,404	11,594	17,216	17,226
#59	147	17,434	11,620	46,338	46,672	11,522	11,562
#60	22,110	11,556	46,367	22,856	22,985	46,396	46,283
#61	6,340	46,624	23,121	7,589	7,770	22,921	22,944
#62	17,458	22,953	7,776	28,793	28,904	7,787	7,739
#63	11,200	7,765	28,649	1,695	1,691	28,730	28,630
#64	47,127	28,957	1,807	3,470	3,574	1,726	1,603
#65	23,200	1,638	3,629	4,242	4,401	3,655	3,620
#66	7,312	3,502	4,355	7,184	7,308	4,228	4,200
#67	29,357	4,275	7,383	21,901	22,087	7,335	7,247
#68	710	7,205	21,853	15,108	15,344	21,988	22,016
#69	3,287	22,040	15,231	2,060	1,955	15,148	15,189
#70	4,260	15,153	2,082	16,491	16,768	2,024	1,860
#71	7,326	2,053	16,649	15,777	15,995	16,699	16,592
#72	22,555	16,701	15,765	4,075	4,164	15,857	15,830
#73	15,334	15,951	4,279	27,577	27,714	4,188	4,234
#74	1,282	4,149	27,370	39,265	39,767	27,895	27,406
#75	16,792	27,656	39,455	11,595	11,744	39,603	39,497
#76	15,963	39,621	11,818	12,390	12,466	11,812	11,770
#77	3,751	11,771	12,321	6,404	6,427	12,451	12,459
#78	28,082	12,444	6,497	18,037	17,965	6,362	6,410
#79	39,969	6,420	17,950	68,620	69,077	18,277	18,223
#80	11,450	18,195	68,107	9,256	9,404	69,288	68,571
#81	12,435	68,527	9,439	6,983	5,266	9,440	9,423
#82	6,117	9,449	7,133	10,492	10,342	4,699	4,729
#83	18,124	7,203	10,372	2,928	2,850	1,423	1,633
#84	69,282	10,364	2,851	8,978	9,120	9,993	9,838
#85	9,181	2,891	9,138	1,817	1,754	2,869	2,852
#86	13	9,026	1,822	8,157	8,202	9,058	9,086
#87	6,697	1,956	8,260	33,064	33,274	1,734	1,727
#88	827	8,265	33,107	5,854	5,811	8,279	8,163
#89	9,999	33,079	5,976	11,489	11,464	33,645	33,311
#90	2,663	5,925	11,625	30,671	30,553	5,939	6,004
#91	8,540	11,544	30,528			11,656	11,670
#92	612	30,784				30,761	30,781

Fragment order number in each profile	**ATCC 23445 *in silico* map**	**ATCC 23445 optical map**	**Bs364CITA optical map**	**Bs396CITA optical map**	**PT09172 optical map**	**PT09143 optical map**	**Bs143CITA optical map**
#93	692						
#94	7,988						
#95	34,269						
#96	5,627						
#97	11,456						
#98	31,209						
Sum of fragments size	1,400,844	1,401,295	1,404,445	1,403,111	1,402,041	1,402,229	1,403,543
Min fragment size	13	1,258	1,487	1,356	1,477	1,336	1,581
Max fragment size	69,282	68,527	68,107	68,620	69,077	69,288	68,571
Average fragment size	14,294	15,231	15,433	15,590	15,578	15,242	15,256
Number of fragments	98	92	91	90	90	92	92

Supplementary Table S3.2. *Brucella* strains used to search for the inversion in chromosome I and the large indel event in the reference strain *B. suis* ATCC 23445.

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
INRA03-26	1	Swine	France	1a	n.a.	995	890	714
AFSSA-03/3081-2	1	Cattle	Croatia	1a	n.a.	995	890	714
AFSSA-04/115	1	Hare	France	1a	n.a.	995	890	714
AFSSA-04/3025-3	1	Swine	Croatia	1a	n.a.	995	890	714
REF 1330 [biovar 1; ATCC 23444]	1	Swine	USA	1a	n.a.	995	890	714
AFSSA-04/2987	1	Hare	France	1a	n.a.	995	890	714
AFSSA-01/5744	1	Swine	Polynesia	1a	n.a.	995	890	714
AFSSA-03/2067-203	1	Swine	France	1a	n.a.	995	890	714
VLA-92/29	1	Hare	Mexico	1a	n.a.	995	890	714
VLA-64/24	1	Swine	USA	1a	n.a.	995	890	714
VLA-F1/04	1	Hare	Netherlands	1a	n.a.	995	890	714
AFSSA-96/1646-01	1	Swine	France	1a	n.a.	995	890	714
REF 686 [biovar 3; ATCC 23446]	3	Swine	USA	3a	n.a.	995	890	714
VLA-63/252	4	Caribou	Alaska	4a	n.a.	995	890	714
REF 40 [biovar 4; ATCC 23447]	4	Reindeer	Former USSR	4a	n.a.	995	890	714
VLA-63/202	4	Reindeer	Poland	4a	n.a.	995	890	714
VLA-63/219	4	Reindeer	Poland	4a	n.a.	995	890	714
VLA-63/198	4	Reindeer	Poland	4a	n.a.	995	890	714
REF 513 [biovar 5; NCTC 11996]	5	Wild rodent	Former USSR	5a	n.a.	995	890	714
AFSSA-92/11580-4528	2	Hare	France	2b	CE	995	890	714
AFSSA-96/9635	2	Swine	France	2c	CE	995	890	714
A183	2	Wild boar	Germany	2a	CE	995	890	714
A196	2	Wild boar	Germany	2a	CE	995	890	714
04RB0377	2	Wild boar	Germany	2a	CE	995	890	714
05RB0007	2	Wild boar	Germany	2a	CE	995	890	714
AFSSA-04/3025-1	2	Swine	Croatia	2c	CE	995	890	714
05RB1442	2	Hare	Germany	2a	CE	995	890	714
REF Thomsen [biovar 1; ATCC 23445]	**2**	**Swine**	**Denmark**	**2a**	**CE**	**995**	**890**	NEG

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
AFSSA-00/9182	2	Hare	France	2b	CE	995	890	714
AFSSA-92/13000	2	Hare	France	2b	CE	995	890	714
Bs364CITA[e]	**2**	**Wild boar**	**Spain**	**2c**	CE	**995**	**890**	714
Bs365CITA	2	Wild boar	Spain	2c	CE	995	890	714
Bs396CITA[e]	**2**	**Wild boar**	**Spain**	**2c**	**CE**	**995**	**890**	714
C9B4	2	Wild boar	Belgium	2c	CE	995	890	714
S275	2	Wild boar	Belgium	2c	CE	995	890	714
C4B6	2	Wild boar	Belgium	2c	CE	995	890	714
AFSSA-04/1918-1	2	Wild boar	Switzerland	2c	CE	995	890	714
C13B4	2	Wild boar	Belgium	2c	CE	995	890	714
C2B11	2	Wild boar	Belgium	2c	CE	995	890	714
C9B3	2	Wild boar	Belgium	2c	CE	995	890	714
PY69	2	Wild boar	Belgium	2c	CE	995	890	714
0111602/4+9	2	Wild boar	Belgium	2c	CE	995	890	714
0111602/3+8	2	Wild boar	Belgium	2c	CE	995	890	714
C8B3	2	Wild boar	Belgium	2c	CE	995	890	714
S120	2	Wild boar	Belgium	2c	CE	995	890	714
C5B5	2	Wild boar	Belgium	2c	CE	995	890	714
AFSSA-03/1483-8	2	Wild boar	France	2c	CE	995	890	714
C13B1	2	Wild boar	Belgium	2c	CE	995	890	714
C3B3	2	Wild boar	Belgium	2c	CE	995	890	714
C13B6	2	Wild boar	Belgium	2c	CE	995	890	714
C11B4	2	Wild boar	Belgium	2c	CE	995	890	714
COSA13	2	Wild boar	Belgium	2c	CE	995	890	714
AFSSA-00/4898	2	Cattle	France	2c	CE	995	890	714
RATES5-11	2	Wild boar	Belgium	2c	CE	995	890	714
C6B1	2	Wild boar	Belgium	2c	CE	995	890	714
It4	2	Wild boar	Italy	2c	CE	995	890	714
It5	2	Wild boar	Italy	2c	CE	995	890	714
It2	2	Wild boar	Italy	2c	CE	995	890	714
It3	2	Wild boar	Italy	2c	CE	995	890	714

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
AFSSA-04/770	2	Wild boar	Italy	2c	CE	995	890	714
It6	2	Wild boar	Italy	2c	CE	995	890	714
It1	2	Wild boar	Italy	2c	CE	995	890	714
AFSSA-97/9757	2	Swine	France	2c	CE	995	890	714
MASA07	2	Wild boar	Belgium	2c	CE	995	890	714
AFSSA-97/4924-10	2	Swine	France	2c	CE	995	890	714
AFSSA-98/7296-4204	2	Hare	France	2a	CE	995	890	714
AFSSA-98/6335	2	Swine	France	2b	CE	995	890	714
LNIV-4193(1)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4193(2)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4193(9)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-22497(475)/08	2	Sheep	Portugal	2d	IB	850	950	714
LNIV-22498(697)/08	2	Sheep	Portugal	2d	IB	850	950	714
LNIV-9H	2	Swine	Portugal	2d	IB	850	950	714
LNIV-44821(81)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-44821(86)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-44821(79)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-3115(9)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-3514(2)/10	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-3515(6)/10	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-44401(3)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-Soc57/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc64/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc73/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc41/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-1344(3)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-14256(72)/09	2	Swine	Portugal	2d	IB	850	950	714
LNIV-231(1)/09	2	Swine	Portugal	2d	IB	850	950	714
LNIV-14256(40)/09	2	Swine	Portugal	2d	IB	850	950	714
LNIV-1989(112)/10	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-3115(13)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4478(4)/09	2	Wild boar	Portugal	2d	IB	850	950	714

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
LNIV-4187(7)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-3478(3)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-468/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc29/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-6552(10)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-2739(3)/02	2	Swine	Portugal	2d	IB	850	950	714
LNIV-7676/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc43/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc35/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc50/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-1262/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-Soc39/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-1967/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-438/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-5346(1)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-5346(16)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-5346(3)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-6552(3)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-6552(8)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-6552(2)/00	2	Swine	Portugal	2d	IB	850	950	714
LNIV-2739(18)/02	2	Swine	Portugal	2d	IB	850	950	714
LNIV-2739(9)/02	2	Swine	Portugal	2d	IB	850	950	714
LNIV-4215(20)/01	2	Swine	Portugal	2d	IB	850	950	714
LNIV-46685(15)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-9789(J10)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-9789/08	2	Wild boar	Portugal	2d	IB	850	950	714
S-32	2	Swine	Spain	2d	IB	850	950	714
LNIV-5414(12)/03	2	Swine	Portugal	2d	IB	850	950	714
LNIV-5414(13)/03	2	Swine	Portugal	2d	IB	850	950	714
LNIV-5414(2)/03	2	Swine	Portugal	2d	IB	850	950	714
LNIV-5414(13)/03	2	Swine	Portugal	2d	IB	850	950	714
S-34	2	Swine	Spain	2d	IB	850	950	714

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
LNIV-44406(3)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-21566(94)/09	2	Wild boar	Portugal	2d	IB	850	950	714
Bs384CITA	2	Swine	Spain	2d	IB	850	950	714
S-12	2	Swine	Spain	2e	IB	850	950	714
LNIV-45014(4)/08	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-4477(4)/09	2	Wild boar	Portugal	2e	IB	850	950	714
S-13	2	Swine	Spain	2e	IB	850	950	714
LNIV-J2A/08	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-9789(J9)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-19122(J1)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-17888(J3)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-21346(J2)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-2454(J10)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-2948(1)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(5)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(6)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(8)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(9)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-1344(2)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4189(1)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-4189(2)/09	2	Wild boar	Portugal	2e	IB	850	950	714
S-22	2	Swine	Spain	2d	IB	850	950	714
Bs147	2	Swine	Spain	2d	IB	850	950	714
Bs146	2	Swine	Spain	2d	IB	850	950	714
LNIV-4498(J1)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4498(J9)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4498(J4)/08	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-4498(J6)/08	2	Wild boar	Portugal	2d	IB	850	950	714
S-3	2	Swine	Spain	2d	IB	850	950	714
S-4	2	Swine	Spain	2d	IB	850	950	714
S-1	2	Swine	Spain	2d	IB	850	950	714
S-2	2	Swine	Spain	2d	IB	850	950	714

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
S-21	2	Swine	Spain	2d	IB	850	950	714
S-6	2	Swine	Spain	2d	IB	850	950	714
LNIV-8605(2)/10	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-8605(9)/10	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-4647(1)/09	2	Wild boar	Portugal	2d	IB	850	950	714
PT09172[e]	**2**	**Wild boar**	**Portugal**	**2d**	**IB**	**850**	**950**	714
LNIV-44821(121)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-44821(122)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-44821(123)/09	2	Wild boar	Portugal	2d	IB	850	950	714
LNIV-46685(20)/08	2	Wild boar	Portugal	2d	IB	850	950	714
Bs144	2	Swine	Spain	2e	IB	850	950	714
Bs145CITA	2	Wild boar	Spain	2e	IB	850	950	714
Bs143CITA[e]	**2**	**Wild boar**	**Spain**	**2e**	**IB**	**850**	**950**	714
LNIV-2948(3)/09	2	Wild boar	Portugal	2e	IB	850	950	714
PT09143[e]	**2**	**Wild boar**	**Portugal**	**2e**	**IB**	**850**	**950**	714
LNIV-2948(12)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(20)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(27)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(29)/09	2	Wild boar	Portugal	2e	IB	850	950	714
LNIV-2948(34)/09	2	Wild boar	Portugal	2e	IB	850	950	714
Bs145	2	Swine	Spain	2e	IB	850	950	714
REF 16M [*B.melitensis* biovar 1; ATCC 23456]	1	Goat	USA	n.a.	n.a.	995	890	714
REF 63/9 [*B.melitensis* biovar 2; ATCC 23457]	2	Goat	Turkey	n.a.	n.a.	995	890	714
REF Ether [*B.melitensis* biovar 3; ATCC 23458]	3	Goat	Italia	n.a.	n.a.	995	890	714
REF 544 [*B.abortus* biovar 1; ATCC 23448]	1	Cattle	England	n.a.	n.a.	995	890	714
2308 [*B.abortus* biovar 1]	1	Cattle	England	n.a.	n.a.	995	890	714
S19 [*B.abortus* biovar 1; vaccine strain]	1	Cattle	England	n.a.	n.a.	995	890	714
REF 86/8/59 [*B.abortus* biovar 2; ATCC 23449]	2	Cattle	England	n.a.	n.a.	995	890	714
REF Tulya [*B.abortus* biovar 3; ATCC 23450]	3	Human	Uganda	n.a.	n.a.	995	890	714
REF 292 [*B.abortus* biovar 4; ATCC 23451]	4	Cattle	USA	n.a.	n.a.	995	890	714

Strain Identification	Biovar	Host species	Country of origin	Haplotype[a]	Clonal lineage[b]	PCR-multiplex[c]		PCR[d]
						Left region (bp)	Right region (bp)	Amplicon (bp)
REF B3196 [*B.abortus* biovar 5; ATCC 23452]	5	Cattle	USA	n.a.	n.a.	995	890	714
REF 870 [*B.abortus* biovar 6; ATCC 23453]	6	Cattle	USA	n.a.	n.a.	995	890	714
REF C68 [*B.abortus* biovar 9; ATCC 23455]	9	Cattle	USA	n.a.	n.a.	995	890	714
REF 63/290 [*B.ovis*; ATCC 25840]	na	Sheep	Australia	n.a.	n.a.	995	890	714
REF B. canis [ATCC 23365]	na	Dog	USA	n.a.	n.a.	995	890	714
REF B2/94 [*B.pinnipedialis*; NCTC 12890]	na	Seal	Scotland	n.a.	n.a.	995	890	714
REF B1/94 [*B.ceti*; NCTC 12891]	na	Dolphin	Scotland	n.a.	n.a.	995	890	714
B. microti strain CCM 4915	na	Wild rodent	Rep.Czech	n.a.	n.a.	995	890	714

[a] PCR-RFLP analysis of *omp31*, *omp2a* and *omp2b* genes using the restriction enzymes AvaII, *Hae*III, *Nco*I, *Sty*I, *Eco*RI and *Kpn*I. P1 to P3 refer to restriction patterns observed for each enzyme; NC: not cleaved.

[b] Clonal lineage of biovar 2 strains: CE, Central-European; IB: Iberian; n.a.: not applicable.

[c] PCR-multiplex at the left and right flanking region to confirm the presence of the genetic inversion in Iberian strains.

[d] PCR to confirm the presence of the indel in *B. suis* strains; NEG, negative in PCR.

[e] Strains used for the production of optical maps.

Chapter 4

Comparative Genomic Analysis of Brucella suis biovar 2 gives support to the distinctiveness of Iberian and Central-European clones

Subchapter 4.1.

Full–genome sequencing of five Brucella suis biovar 2 strains representative of the two circulating clonal lineages in Iberian Peninsula

The results presented in this subchapter were previously published.

Ferreira A.C., Tenreiro R., Corrêa de Sá and M.I. Dias R. (2014). Complete genome sequences of three Iberian *Brucella suis* biovar 2 strains isolated from wild boars. *Genome Announcements.* 2(4):e00618-14. doi:10.1128/genomeA.00618-14.

Ferreira A.C., Tenreiro R., Corrêa de Sá and M.I. Dias R. (2014). Complete genome sequences of two Central-European *Brucella suis* bv. 2 haplotype 2c strains isolated from wild boars. *Genome Announcements***.** 2(4):e00686-14. doi:10.1128/genomeA.00686-14.

Authors contributions:

ACF and **RD** carried out the *de novo* sequencing and assembly, the functional annotation and drafted the manuscript.

MICS, RT and **RD** conceived of the study, participated in its design and coordination, and revised the manuscript.

1. Introduction

Brucella suis biovar 2 infection in wild boar (*Sus scrofa*) is widespread in the Iberian Peninsula, and it was previously described that the majority of *B. suis* biovar 2 strains circulating in Portugal and Spain share specific molecular characteristics establishing an Iberian clonal lineage (Ferreira *et al.* 2016). Yet, above Ebro's river, in the North-East region of Spain, it has been isolated strains from wild boars similar to those isolated in different Central-European countries, in both pigs and wild boars (Muñoz *et al.*, 2010).

Knowing the genome as exactly as possible is of fundamental value to microbial biology. In the past thirty years, DNA sequencing technologies and applications have undergone tremendous development, driven the development of second-generation sequencing methods, or next generation sequencing (NGS), that became an alternative to traditional Sanger DNA sequencing (Grada and Weinbrecht, 2013). The Illumina next-generation sequencing technology (www.illumina.com) is a breakthrough platform based on massively parallel sequencing of millions of fragments that uses its proprietary reversible terminator-based sequencing chemistry. This approach ensures high accuracy and true base-by-base sequencing, eliminating sequence-context specific errors and enabling sequencing through homopolymers and repetitive sequences. Sequence reads are aligned against a reference genome and genetic differences are called using specially developed data analysis pipeline software. The Illumina sequencing workflow is based on three steps: libraries preparation from any nucleic acid sample, production of clonal clusters by amplification, and sequencing using massively parallel synthesis. Template preparation consists of building a library of nucleic acids (DNA or complementary DNA - cDNA) and amplifying that library. Sequencing libraries are constructed by fragmenting the DNA (or cDNA) sample and ligating adapter sequences (synthetic oligonucleotides of a known sequence) onto the ends of the DNA fragments. Once constructed, libraries are clonally amplified in preparation for sequencing. Illumina utilizes ridge amplification to form template clusters on a flow cell (Quail *et al.*, 2012; Berglund *et al.*, 2011).

Whole-genome Optical Mapping (WGM) technology is a recent tool that creates a single molecule high-resolution *in situ* ordered restriction map of a bacterial genome (Shukla *et al.*, 2009). Such optical maps were used with success not only to determine genomic organization and perform comparative genomics to identify genomic rearrangements (such as insertions, deletions, duplications, and inversions), but also to validate and finish sequence contigs and genomic placement during sequencing projects (Ferreira *et al.*, 2016; Mariano *et al.*, 2016; Bosch *et al.*, 2013; Hall *et al.*, 2013; Fey *et al.*, 2012; Mellmann *et al.* 2011; Deurenberg *et al.*, 2007; Kotewicz *et al.*, 2007; Valouev *et al.*, 2006; Zhou *et al.*, 2004; Lim *et al.*, 2001).

Consensus optical maps of five *B. suis* biovar 2 field strains isolated from wild boars in Portugal and Spain were previously generated from the restriction patterns obtained at single-molecule level with *BamH* I (Ferreira *et al.*, 2016). In this work the genomic sequences of these *B. suis* biovar 2 strains, including three Iberian (one from haplotype 2d and two haplotype 2e) and two Central-European (haplotype 2c), were obtained by Illumina HiSeq 2000 technology using a paired-end 35 bp protocol, reads were *de novo* assembled using de Bruijn graph method, and scaffolding was guided by optical mapping method, allowing high confidence and prompt disclosure of typical sequence assembly problems.

2. Material and Methods

2.1. *B. suis* strains and DNA isolation

From the set of biovar 2 isolates, five strains representative of the two circulating clonal lineages in Iberian Peninsula were chosen and sequenced to full closure. *B. suis* biovar 2 strains PT09143 (haplotype 2e) and PT09172 (haplotype 2d) were isolated from wild boars in two different regions of Portugal, one in the south (Alentejo) and other in the North (Trás-os-Montes), respectively. Strains Bs143CITA (haplotype 2e), Bs364CITA and Bs396CITA (both from haplotype 2c) were kindly provided by JM Blasco from the Centro de Investigación y Tecnología Agroalimentaria de Aragón (CITA, Zaragoza, Spain). The three strains were isolated from wild boars from two regions in the North of Spain, Asturias (Bs143CITA) and Aragón (northeast region, above the Ebro River) (**Table 4.1.1**). Strains were previously biotyped according to standard bacteriological procedures (Alton *et al.*, 1988), and characterized through PCR-restriction fragment length polymorphism (RFLP) analysis for the *omp2a*, *omp2b*, and *omp31* genes (Vizcaíno *et al.*, 1997; Cloeckaert *et al.*, 1995) and by multi*locus* variable-number tandem-repeat (VNTR) assay (MLVA) (Le Flèche *et al.*, 2006) as described before (see **Supplementary Tables S2.2.1 and S2.2.2 in Subchapter 2.2).** Total genomic DNA was extracted and purified using the PureLink Genomic DNA Kit (Invitrogen Life Technologies, USA). An aliquot of the DNA was subjected for analysis using the Bioanalyzer (Agilent technologies) and was confirmed for no degradation. An aliquot of 10 µg of DNA was used for the sequencing and the remaining stock was maintained for further sequencing and completion of gaps. In addition, reference strains *B. suis* 1330 (biovar 1; ATCC 23444) and *B. suis* strain Thomsen (biovar 2; ATCC 23445) were obtained from the American Type Culture Collection (ATCC, USA), and used for validation of PCR methods.

2.2. Sequencing and *de novo* assembly

Genomic libraries were created using Truseq DNA sample preparation kit and genomic sequences obtained by Illumina HiSeq 2000 technology with a paired-end 35-bp protocol. The quality analysis of the reads was done using the FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Only reads with Phred score Q>30 (*i.e.*, Error Probability <0.001) were considered for *de novo* assembly (depth coverage aptx: 100X), using de Bruijn graph method (Velvet version 1.2.09; Zerbino and Birney, 2008). Visual inspection of alignment of fastq reads for each contig was performed using Table vs 1.14.04.10. Scaffolding was guided by optical mapping method (MapSolver version 3.2; OpGen Technologies, Inc.) and gap filling was performed by PCR and Sanger method (Sanger *et al.*, 1977). Low coverage regions and all regions containing insertions or deletions (INDELs) in comparison with *B. suis* ATCC 23445, were confirmed by Sanger resequencing (**Supplementary Table S4.1.1**). Chromosome-wide comparisons of those five strains and *B. suis* reference strains ATCC 23445 and 1330 were made using an approach based in Mummer algorithm implemented at Kodon V3.62. Genomic alignment of concatenated chromosome was performed by superstretch approach: DNA seed 15 matches in windows size of 25 bases, minimal stretch length 60 bases, and minimal cut-off for stretch identity of 60% in a screening window of 30 bases were used.

2.3. Sequence-to-map comparison

Genomic DNA from the five strains was extracted and optical maps were prepared using the Argus Whole Genome Mapping System (OpGen, Inc., Madison, USA). A consensus optical map was produced for each strain from the restriction pattern obtained at single-molecule level. Sequence-to-map comparisons were performed using MapSolver software (OpGen Technologies, Inc.). Sequence FASTA files from each chromosome were converted to *in silico* restriction maps for direct comparison to the Optical maps. Comparisons were accomplished by aligning the sequence with the optical maps according to their restriction fragment pattern. Alignments were generated with a dynamic programming algorithm which finds the optimal location, or placement, of a sequence contig by first performing a global alignment of the sequence contig against the optical map.

2.4. Functional annotation

The ORF detection and primary functional annotation was made through Rapid Annotations using Subsystems Technology (RAST) (Aziz *et al.*, 2008), 23S rRNA and tRNA genes were identified using RNAmmer (Lagesen *et al.*, 2007) and tRNAscan-SE 1.21 (Lowe and Eddy, 1997). Functional enrichment was performed using Gene Ontology (GO) (The Gene Ontology Consortium, 2000), InterPro and Kyoto

Encyclopedia of Genes and Genomes (KEGG) databases, using Blast2GO pipeline (version 2.7.1) (Götz *et al.*, 2008). Search for phage at the sequenced genomes and strains *B. suis* ATCC 23445 and 1330 was made using PHAST (PHAge Search Tool) (Zhou *et al.*, 2011). For ORFs without functional information attributed by RAST server nor Blast2GO, annotation was obtained by query with BLASTn algorithm against nr/nt database from NCBI (May 10, 2014) using standard configuration. Annotation was curated manually. Data consolidation, mining and querying was performed using MySQL InnoDB engine (5.0.95). Scripting and parsing was made using Python (2.7.3) (Van Rossum, 2007).

2.5. Nucleotide sequence accession numbers

Genome accession numbers to GenBank database are listed in **Table 4.1.1.**

Table 4.1.1. Listing of strains and respective genome accession numbers.

Genome	Clonal lineage (Haplotype)	Country	GenBank accession number (chr I/ chrII)
PT09172	Iberian (2d)	Portugal	CP007693.1/ CP007694.1
PT09143	Iberian (2e)	Portugal	CP007691.1/ CP007692.1
Bs143CITA	Iberian (2e)	Spain	CP007695.1/ CP007696.1
Bs364CITA	Central-European (2c)	Spain	CP007697.1/ CP007698.1
Bs396CITA	Central-European (2c)	Spain	CP007720.1/ CP007721.1

3. Results and Discussion

The full genomes of PT09172, PT09143, Bs143CITA, Bs364CITA and Bs396CITA were reconstructed from a range of 8,408,102 to 12,172,794 high-quality reads (Phred score >30) per strain, generated by Illumina HiSeq 2000 technology, with a paired-end 35-bp protocol. Reads were *de novo* assembled into contigs using de Bruijn graph method, resulting in depth coverage ranging from 89X to 128X. The assembly yielded between 55 to 116 contigs and the genome expected sizes ranged from 3,293,158 to 3,304,238 bp. Summary of statistics of assembly process is presented in **Table 4.1.2.**

For all strains, the structural chromosomal organization was confirmed using high-resolution whole-genome optical mapping (OpGen Technologies, Inc.). Optical mapping produces a barcode-like genetic map of restriction sites with the depicted sites arranged in the order they occur in the genome. The consensus map was built within a minimal coverage of 50X, generating 228 to 232 fragments, with sizes ranging from 1.5 to 85 kb (see **Supplementary Tables S3.2A e S3.2B** in Chapter 3). The optical maps defined unique genome landmarks in each of the strains and demonstrated the ability of this

method to determine the relative placement and orientation of sequence fragments produced during the assembly process.

4.1.2. Summary statistics for assembly of five *B. suis* biovar 2 strains isolates from wild boars.

Statistics	**PT09143**	**PT09172**	**Bs143CITA**	**Bs364CITA**	**Bs396CITA**
Number of raw reads	8,408,102	12,172,794	10,795,456	10,091,666	8,423,200
Sequencing Coverage depth (X)	89.36	128.13	113.65	106.10	88.57
Number of reads used	8,123,604	11,727,505	10,434,053	9,524,505	8,150,997
Kmer used (Velvet)	21	23	21	29	21
Estimated Genome Size (bp)	3,293,158	3,297,697	3,301,707	3,302,044	3,304,238
Number of molecules	2	2	2	2	2
Median coverage depth	35.59	44.86	45.90	19.79	35.70
Number of contigs	116	65	61	55	77
N50	56,902	100,353	122,343	123,75	83,863
Max. contig (bp)	163,9	208,981	284,642	256,126	215,336
Min. contig (bp)	611	611	611	611	542

General features of the *B. suis* biovar 2 genomes are summarized in **Table 4.1.3**. The five *B. suis* biovar 2 genomes have similar sizes, and are composed of two circular chromosomes (I and II) with approximately 1.93 and 1.40 Mb, and the genome sequence length ranged from 3,324,539 (Bs143CITA) and 3,328,972 bp (Bs364CITA). The overall GC contents of Chr I and Chr II were 57.1% and 57.3%, respectively. The five genomes contain three identical rRNA operons (**Supplementary Table S4.1.2**), one in Chr I and two located in Chr II, and a set of 54 tRNAs genes were predicted (**Supplementary Table S4.1.3**). Established from homology-based search and from protein-structural-motif identification, high-confidence protein-coding genes were predicted ranging from 3,014 (PT09143) to 3,027 (Bs396CITA) coding sequences, and the number of pseudogenes varied between 87 (Bs396CITA) and 91 (Bs364CITA). The predicted coding sequences (CDS) and respective functional identification for each genome can be seen in **Supplementary Table S4.1.4,** allocated in http://www.mediafire.com/file/ducz93rz73wmu29/Supplementary_Table_S4.1.4_A-E.xlsx. Both chromosomes are highly similar to the two chromosomes of *B. suis* biovar 2 ATCC 23445 reference strain. However, unexpectedly, the combined use of NGS and optical mapping technologies permitted the identification and location of a genetic inversion of ~944 kb in Chr I of the three strains from Iberian clonal lineage. This inversion has been demonstrated to be specific of Iberian clonal lineage strains (haplotype 2d and 2e) (Ferreira *et al.*, 2016).

Table 4.1.3. Characteristics of *Brucella suis* biovar 2 genomes.

Genome	PT09143	PT09172	Bs143CITA	Bs364CITA	Bs396CITA
Chr I size (bp)	1,926,480	1,926,716	1,926,295	1,927,594	1,927,083
Chr II size (bp)	1,398,285	1,398,326	1,398,244	1,401,378	1,401,375
Total size (bp)	3,324,765	3,325,042	3,324,539	3,328,972	3,328,458
CDS	3,014	3,015	3,015	3,018	3,027
Genes	3,167	3,168	3,168	3,173	3,178
Hypothetical proteins	774	726	790	750	745
Pseudogenes	89	89	89	91	87
Frameshifted genes	67	65	67	66	62
rRNA	9	9	9	9	9
tRNA	54	54	54	54	54
ncRNA	1	1	1	1	1

Chr, chromosome; CDS, coding sequence; rRNA, ribosomal RNA; tRNA, transfer RNA; ncRNA, non-coding RNA

4. Conclusion

The full genomes of the five *B. suis* biovar 2 field strains were promptly obtained using a combination of Illumina, Sanger and optical mapping technologies. Until now, only the whole-genome of the reference strain of *B. suis* biovar 2 strain Thomsen (ATCC 23445) was available for comparative analysis. The release of additional complete genomes of other strains isolated from different animal hosts may facilitate intra-generic comparative analysis of the species or genus. Further, as more whole-genome sequences of different strains are published, the biological implications of genomic plasticity in *Brucella* phylogeny will be clarified.

5. Supplementary material

Supplementary Table S4.1.1. Low coverage regions and indels confirmed by Sanger resequencing (Reference positions to *B. suis* ATCC 23445).

Supplementary Table S4.1.2. Ribosomal RNA (rRNA) operons identified using RNAmmer.

Supplementary Table S4.1.3. Transference RNA (tRNA) genes identified using tRNAscan-SE 1.21.

Supplementary Table S4.1.4. Predicted coding sequences (CDS) and respective functional annotation for *B. suis* biovar 2 genomes. Data allocated at:

http://www.mediafire.com/file/ducz93rz73wmu29/Supplementary_Table_S4.1.4_A-E.xlsx

6. References

1. **Alton, G.C., Jones, L.M., Angus, R.D. & Verger, J.M. (1988**). Techniques for the Brucellosis Laboratory. Institut National de la Recherche Agronomique (INRA), Paris, France.
2. **Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. (2008)**. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75.
3. **Berglund, E.C., Kiialainen, A. & Syvänen, A.C. (2011)**. Next-generation sequencing technologies and applications for human genetic history and forensics. Investigative Genetics, 2: 23. doi: 10.1186/2041-2223-2-23.
4. **Bosch, T., Verkade, E., van Luit, M., Pot, B., Vauterin, P., Burggrave, R., Savelkoul, P., Kluytmans, J. & Schouls, L. (2013)**. High resolution typing by whole genome mapping enables discrimination of LA-MRSA (CC398) strains and identification of transmission events. PLoS ONE, 8(6): e66493. doi:10.1371/journal.pone.0066493. doi:10.1371/journal.pone.0066493.
5. **Cloeckaert, A., Verger, J.M., Grayon, M. & Grépinet, O. (1995)**. Restriction site polymorphism of the genes encoding the major 25 and 36 kDa outer-membrane proteins of *Brucella*. Microbiology, 141: 2111–2121.
6. **Ferreira, A.C., Dias. R., de Sá. M.I. & Tenreiro, R. (2016)**. Whole-genome mapping reveals a large chromosomal inversion on Iberian *Brucella suis* biovar 2 strains. Veterinary Microbiology, 192:220-5.
7. **Fey, P.D., Iwen, P.C., Zentz, E.B., Briska, A.M., Henkhaus, J.K., Bryant, K.A., Larson, M.A., Noel, R.K. & Hinrichs, S.H. (2012)**. Assessment of whole-genome mapping in a well-defined outbreak of Salmonella enterica serotype Saintpaul. Journal of Clinical Microbiology, 50(9): 3063-5. doi: 10.1128/JCM.01320-12.
8. **Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium (2000).** Nature Genet. 25: 25-29.
9. **Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J. & Conesa, A. (2008)**. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res.36(10): 3420-35.
10. **Grada, A. & Weinbrecht, k. (2013)**. Next-Generation Sequencing: Methodology and Application. Journal of Investigative Dermatology, 133: e11. doi:10.1038/jid.2013.248.
11. **Hall, B.G., Kirkup, B.C., Riley, M.C. & Barlow, M. (2013)**. Clustering acinetobacter strains by optical mapping. Genome Biology and Evolution, 5(6): 1176-84. doi: 10.1093/gbe/evt085.
12. **Kotewicz, M.L., Jackson, S.A., LeClerc, J.E. & Cebula, T.A. (2007)**. Optical maps distinguish individual strains of *Escherichia coli* O157: H7. Microbiology, 153: 1720-1733.
13. **Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T. & Ussery, D.W. (2007)**. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100 –3108.
14. **Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoed, P., Nöckler, K., Neubauer, H., Guilloteau, L.A. & Vergnaud, G. (2006)**. Evaluation and selection of tandem repeat *loci* for a Brucella MLVA typing assay. BMC Microbiology, 6:9-22.
15. **Lim, A., Dimalanta, E. T., Potamousis, K.D., Yen, G., Apodoca, J., Tao, C., Lin, J., Qi, R., Skiadas, J., Ramanathan, A., Perna, N.T., Plunkett, G. 3rd, Burland, V., Mau, B., Hackett, J., Blattner, F.R., Anantharaman, T.S., Mishra, B. & Schwartz, D.C. (2001)**. Shotgun Optical Maps of the Whole *Escherichia coli* O157:H7 Genome. Genome Research, 11(9): 1584–1593. http://doi.org/10.1101/gr.172101.
16. **Lowe, T.M. & Eddy, S.R. (1997)**. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.
17. **Mariano, D.C., Sousa Tde, J., Pereira, F.L., Aburjaile, F., Barh, D., Rocha, F., Pinto, A.C., Hassan, S.S., Saraiva, T.D., Dorella, F.A., de Carvalho, A.F., Leal, *C.A.*, Figueiredo, H.C., Silva, A., Ramos, R.T. & Azevedo, V.A. (2016)**. Whole-genome optical mapping reveals a mis-assembly between two rRNA operons of *Corynebacterium pseudotuberculosis* strain 1002. BMC Genomics., 17:315.
18. **Mellmann A, Harmsen D, Cummings CA, Zentz EB, Leopold SR, Rico A, Prior K, Szczepanowski R, Ji Y, Zhang W, McLaughlin SF, Henkhaus JK, Leopold B, Bielaszewska M, Prager R, Brzoska PM, Moore RL, Guenther**

S, Rothberg JM, Karch H. (2011). Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One, 6(7): e22751. doi: 10.1371/journal.pone.0022751.

19. **Miller, J.M. (2013)**. Whole-genome mapping: a new paradigm in strain-typing technology. Journal of Clinical Microbiology, 51(4): 1066-70. doi: 10.1128/JCM.00093-13.
20. **Muñoz, P.M., Boadella, M., Arnal, M., de Miguel, M.J., Revilla, M., Martínez, D., Vicente, J., Acevedo, P., Oleaga, A., Ruiz-Fons, F., Marín, C.M., Prieto, J.M., de la Fuente, J., Barral, M., Barberán, M., de Luco, D.F., Blasco, J.M. & Gortázar, C. (2010)**. Spatial distribution and risk factors of brucellosis in Iberian wild ungulates. BMC Infectious Diseases, 10: 46.
21. **Petkau, A., Stuart-Edwards, M., Stothard, P. & Van Domselaar, G. (2010)**. Interactive microbial genome visualization with GView. Bioinformatics 26: 3125-3126.
22. **Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. & Gu, Y. (2012)**. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics, 13: 341. doi: 10.1186/1471-2164-13-341.
23. **Sanger, F., Nicklen, S. & Coulson, A.R. (1977)**. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences of the United States of America, 74:5463–5467.
24. **Valouev, A., Zhang, Y., Schwartz, D.C. & Waterman, M.S. (2006)**. Refinement of optical map assemblies. Bioinformatics, 22(10): 1217-24.
25. **Van Rossum, G. (2007)**. Python programming language. In USENIX Annual Technical Conference.
26. **Vizcaíno, N., Verger, J.M., Grayon, M., Zygmunt, M.S. & Cloeckaert, A. (1997)**. DNA polymorphism at the omp-31 *locus* of *Brucella* spp.: evidence for a large deletion in *Brucella abortus*, and other species-specific markers. Microbiology, 143: 2913–2921.
27. **Zerbino, D.R. & Birney, E. (2008)**. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18:821– 829.
28. **Zhou, Y., Liang, Y., Lynch, K., Dennis, J.J. & Wishart, D.S. (2011)**. PHAST: A Fast Phage Search Tool. Nucl. Acids Res. 39(suppl 2): W347-W352 [doi:10.1093/nar/gkr485.
29. **Zhou, S., Kile, A., Bechner, M., Place, M., Kvikstad, E., Deng, W., Wei, J., Severin, J., Runnheim, R., Churas, C., Forrest, D., Dimalanta, E.T., Lamers, C., Burland, V., Blattner, F.R. & Schwartz, D.C. (2004)**. Single-molecule approach to bacterial genomic comparisons via optical mapping. Journal of Bacteriology, 186(22): 7773-82.

Supplementary Table S4.1.1. Low coverage regions and indels confirmed by Sanger resequencing (Reference positions to *B. suis* ATCC 23445).

Locus	Sequence (5`- 3`)	Chr.	Start	Stop	Product (bp)	Observation
Bs_I-63691F	CTCACGGCAGAAAAGAAAGG	I	63,691	63,952	**262**	**Low coverage region**
Bs_I-63691R	AAATATCCGCGTCGAGAATG					
Bs_I-79518F	TGCACACGTAGGGTCGATA	I	79,420	80,357	**938**	**INDEL**
Bs_I-79518R	CACCCAGATGTTCGGCTAT					
Bs-I_125655F	CGGATTCCTCTTCACCTTCA	I	134,207	134,309	**103**	**INDEL**
Bs_I-125655R	GGGCAGCTCATGACTTTTGT					
Bs_I-143457F	GTCATCACGCTCCAGGTCTT	I	143,472	----------	**No amplification In ATCC 23445**	**INDEL**
BS-I_143457R	CCTGCACACATCAGAACGTC					
Bs_I-649504F	AGTTGAAGCGCAGCTGAGTA	I	649,504	650,234	**731**	**Low coverage region**
Bs_I-649504R	TGCTAAATTGTGGGCCTTTT					
Bs_I-650861F	CCGAAGCCGTTGAATATGT	I	650,770	651,020	**251**	**INDEL**
Bs_I-650861R	CAGCATAGTTGAAGCGCAG					
Bs_I-892731F	ACGCCCTTGATGGTGATTA	I	892,313	892,927	**615**	**INDEL**
Bs_I-892731R	TTCGATTCTGACGTGACCTG					
Bs_I-1022840F	GCTGTGACGGGTAATTTGAA	I	1,022,785	1,023,071	**287**	**INDEL**
Bs_I-1022840R	GTGCTTGCCAATGAGAAGAC					
Bs_I-1036525F	CCAGAAATCAGTCCAGTGGA	I	1,036,480	1,036,679	**200**	**INDEL**
Bs_I-1036525R	GAGCTTGTGTTTGGTGGTGT					
Bs_I-1042003F	ACTGAACAGGCTGAGAGCC	I	1,041,944	1,042,224	**281**	**INDEL**
Bs_I-1042003R	GGAATAATTTTCCACGCCA					
Bs-I_1115894F	TGCGGTCATGAATGTGTTCT	I	1,115,794	1,116,428	**635**	**INDEL**
Bs-I_1115894R	CAGCGAGGATTACGATCCAT					
Bs_I-1171662F	AAGATGTGGATCCGGTGAC	I	1,171,597	1,171,839	**243**	**INDEL**
Bs_I-1171662R	AAGTTATTGCGCTTCAGCG					
Bs_I-1207470F	CCGGTGTTATTTTCGTTTCA	I	1,207,421	1,207,590	**170**	**INDEL**
Bs_I-1207470R	AAAATATCGAAGACCTCGCC					
Bs_I-1355983F	TAGGCTGCCTGGAATTCATC	I	1,355,304	1,356,288	**985**	**INDEL**
Bs_I-1355983R	CGGCAAGTTCACCTCTGACT					
Bs_I-1420091F	GCATATTTTACCCCAAAGCC	I	1,420,091	1,420,711	**621**	**Low coverage region**
Bs_I-1420091R	ATTTTGAACCAGAGCGGTTG					
Bs_I-1423445F	TTTCTTATTCCACCCGATCC	I	1,423,358	1423655	**298**	**Low coverage region**
Bs_I-1423445R	CAAAATGTTACTGCGTGAAGC					
Bs_I-1439474F	CGCGTTCAGAATAGTTCAGTG	I	1,439,402	1,439,604	**203**	**INDEL**
Bs_I-1439474R	GAACGTCCATAGGTCCGAA					
Bs_I-1627483F	CAATCCACAGGAGATCGGT	I	1,627,421	1,627,602	**182**	**INDEL**
Bs_I-1627483R	GCTTCACGGGTATCATGTACA					
Bs_I-1660136F	TTCTCTTCCGTGCTGCCTAT	I	1,660,136	1,660,747	**612**	**Low coverage region**
Bs_I-1660136R	GGAGGGTGACAGAAGAATGG					
Bs_I-1902495F	ATGAGAAGCCGGAAGTTGA	I	1,902,416	1,903,238	**823**	**INDEL**
Bs_I-1902495R	CTTCGCCCTTCACAACCTT					
Bs_II-84685F	ACTGGCGATGATCGAAAAG	II	84,616	84,868	**253**	**INDEL**
Bs_II-84685R	GCTGCAACCATGGTTTCAC					
Bs_II-84796F	ACGAGGAAGAGGATAATGGC	II	84,732	84,848	**117**	**INDEL**
Bs_II-84796R	GCTGCAACCATGGTTTCAC					

Locus	Sequence (5`- 3`)	Chr.	Start	Stop	Product (bp)	Observation
Bs_II-117281F	TCGGAAATGGACGAATATCA	**II**	117,306	117,533	**228**	**INDEL**
Bs_II-117281R	TTCTTGTCGTCGGAAATGTC					
Bs_II-350983F	CCATTGGTTTGAAGCGTTCT	**II**	350,806	351,225	**420**	**INDEL**
Bs_II-350983R	CAGCACATTCGAGGATGAGA					
Bs_II-524675F	ACCCGTCGTCTACGACAATC	**II**	524,559	525,024	**466**	**INDEL**
Bs_II-524675R	ATAACCTTCGGTCGGAACA					
Bs_II-679647F	AGTGGATTTTGGTGCGTTTC	**II**	679,366	680,717	**1,352**	**INDEL**
Bs_II-679647R	AAGATGAGCGGGAAATGTTG					
Bs_II-680354F	CTTTATTCGCAGCGGTAGG	**II**	680,064	680,542	**479**	**INDEL**
Bs_II-680354R	CCGACATATTTGTTGAACGAA					
Bs_II-779882F	GCCAAGTGTGACCGGTATTA	**II**	779,838	780,069	**232**	**INDEL**
Bs_II-779882R	ACGGATTGTCCCTCTCCTT					
Bs_II-799189F	AGCTGCGCGAAATAAAAATC	**II**	799,189	799,841	**653**	**Low coverage**
Bs_II-799189R	AGATTGCAGGCCATCTTCAT					**region**
Bs_II-1098566F	CACCCCCAACTACTCACAAA	**II**	1,098,538	1,098,776	**239**	**INDEL**
Bs_II-1098566R	TCTCGTTTTACGGCGGAAT					
Bs_II-1146197F	GTAACGGCCCTGACCATAGA	**II**	1,145,848	1,146,540	**693**	**INDEL**
Bs_II-1146197R	CGGTTTGAGGAACTGGAAAA					
Bs_II-1303711F	CACCCCCAACTACTCACAAA	**II**	1,303,711	1,304,495	**785**	**Low coverage**
Bs_II-1303711R	CGCTTCTGTGAGTTCTGTGC					**region**
Bs_II-1310854F	TATTATTCACTTTGAGCGGCA	**II**	1,310,801	1,312,970	**2170**	**INDEL**
Bs_II-1310854R	AACTGCAAAAGCTTGGCTG					

F, primer forward; R, primer reverse; I, chromosome I; II, chromosome II

Supplementary Table S4.1.2. Ribosomal RNA (rRNA) operons identified using RNAmmer.

rRNA	Chr.	PT09143	PT09172	Bs143CITA	Bs364CITA	Bs396CITA
5S	I	1,600,883..1,600,996	1,601,095..1,601,208	1,600,779..1,600,892	1,601,977..1,602,090	1,601,441..1,601,554
23S	I	1,601,212..1,604,120	1,601,424..1,604,332	1,601,108..1,604,016	1,602,306..1,605,214	1,601,770..1,604,678
16S	I	1,604,894..1,606,366	1,605,106..1,606,578	1,604,790..1,606,262	1,605,988..1,607,460	1,605,452..1,606,924
5S	II	2,989,860..2,989,973	2,990,105..2,990,218	2,989,633..2,989,746	2,991,970..2,992,083	2,991,453..2,991,566
23S	II	2,990,189..2,993,097	2,990,434..2,993,342	2,989,962..2,992,870	2,992,299..2,995,207	2,991,782..2,994,690
16S	II	2,993,871..2,995,343	2,994,116..2,995,588	2,993,644..2,995,116	2,995,981..2,997,453	2,995,464..2,996,936
5S	II	3,195,048..3,195,161	3,195,294..3,195,407	3,194,819..3,194,932	3,197,154..3,197,267	3,196,641..3,196,754
23S	II	3,195,377..3,198,285	3,195,623..3,198,531	3,195,148..3,198,056	3,197,483..3,200,391	3,196,970..3,199,878
16S	II	3,199,059..3,200,531	3,199,305..3,200,777	3,198,830..3,200,302	3,201,165..3,202,637	3,200,652..3,202,124

I, chromosome I; II, chromosome II

Supplementary Table S4.1.3. Transference RNA (tRNA) genes identified using tRNAscan-SE 1.21.

tRNA Type	Chr.	Anti-Codon	PT09143			PT09172			Bs143CITA			Bs364CITA			Bs396CITA		
			Begin	End	size	Begin	End	size	Begin	End	size	Begin	End	size	Begin	End	size
Ala	I	GGC	21,049	20,977	-72	21,049	20,977	-72	21,049	20,977	-72	21,049	20,977	-72	21,049	20,977	-72
Pro	I	CGG	200,015	200,088	73	115,315	115,388	73	200,019	200,092	73	200,692	200,765	73	200,692	200,765	73
Phe	I	GAA	272,278	272,206	-72	170,643	170,571	-72	272,283	272,211	-72	272,955	272,883	-72	272,955	272,883	-72
Thr	I	GGT	278,565	278,636	71	176,930	177,001	71	278,570	278,641	71	279,242	279,313	71	279,242	279,313	71
Arg	I	ACG	292,644	292,571	-73	191,009	190,936	-73	292,649	292,576	-73	293,310	293,237	-73	293,310	293,237	-73
Glu	I	TTC	421,393	421,322	-71	421,373	421,302	-71	421,396	421,325	-71	527,445	527,516	71	527,389	527,460	71
Tyr	I	GTA	442,197	442,278	81	442,178	442,259	81	442,201	442,282	81	646,470	646,383	-87	646,415	646,328	-87
Gly	I	TCC	442,306	442,376	70	442,287	442,357	70	442,310	442,380	70	720,407	720,480	73	720,296	720,369	73
Trp	I	CCA	443,802	443,874	72	443,783	443,855	72	443,806	443,878	72	721,045	721,118	73	720,934	721,007	73
Leu	I	GAG	607,161	607,080	-81	607,085	607,004	-81	607,178	607,097	-81	779,682	779,609	-73	779,612	779,539	-73
Leu	I	GAG	615,417	615,336	-81	615,341	615,260	-81	615,434	615,353	-81	794,139	794,067	-72	794,070	793,998	-72
Leu	I	TAA	643,998	643,916	-82	643,921	643,839	-82	644,003	643,921	-82	794,549	794,622	73	794,480	794,553	73
Gly	I	CCC	725,659	725,729	70	725,644	725,714	70	725,652	725,722	70	887,637	887,716	79	887,245	887,324	79
Arg	I	CCG	726,510	726,437	-73	726,495	726,422	-73	726,503	726,430	-73	926,681	926,609	-72	926,292	926,220	-72
Val	I	CAC	726,927	726,998	71	726,912	726,983	71	726,920	726,991	71	952,585	952,514	-71	952,196	952,125	-71
Thr	I	TGT	752,832	752,904	72	752,816	752,888	72	752,825	752,897	72	953,002	953,075	73	952,613	952,686	73
Leu	I	TAG	791,876	791,797	-79	791,867	791,788	-79	791,872	791,793	-79	953,853	953,783	-70	953,464	953,394	-70
Asp	I	GTC	884,821	884,748	-73	884,812	884,739	-73	884,818	884,745	-73	1,035,565	1,035,647	82	1,035,175	1,035,257	82
Val	I	TAC	885,231	885,303	72	885,222	885,294	72	885,228	885,300	72	1,064,143	1,064,224	81	1,063,753	1,063,834	81
Pro	I	GGG	899,689	899,762	73	899,680	899,753	73	899,686	899,759	73	1,072,399	1,072,480	81	1,072,009	1,072,090	81
Arg	I	TCT	958,312	958,239	-73	958,455	958,382	-73	958,316	958,243	-73	1,235,772	1,235,700	-72	1,235,382	1,235,310	-72
Pro	I	TGG	958,950	958,877	-73	959,093	959,020	-73	958,954	958,881	-73	1,237,268	1,237,198	-70	1,236,898	1,236,828	-70
Ser	I	GCT	1,032,887	1,032,974	87	1,033,029	1,033,116	87	1,032,885	1,032,972	87	1,237,377	1,237,296	-81	1,237,007	1,236,926	-81
Gln	I	TTG	1,151,912	1,151,841	-71	1,152,053	1,151,982	-71	1,151,908	1,151,837	-71	1,258,190	1,258,261	71	1,257,812	1,257,883	71

tRNA Type	Chr.	Anti-Codon	PT09143			PT09172			Bs143CITA			Bs364CITA			Bs396CITA		
			Begin	End	size	Begin	End	size	Begin	End	size	Begin	End	size	Begin	End	size
Lys	I	CTT	1,313,446	1,313,518	72	1,313,485	1,313,557	72	1,313,345	1,313,417	72	1,314,327	1,314,399	72	1,313,950	1,314,022	72
Met	I	CAT	1,445,417	1,445,344	-73	1,445,639	1,445,566	-73	1,445,315	1,445,242	-73	1,446,532	1,446,459	-73	1,445,968	1,445,895	-73
Met	I	CAT	1,563,802	1,607,730	73	1,564,015	1,564,088	73	1,563,696	1,563,769	73	1,564,892	1,564,965	73	1,564,355	1,564,428	73
His	I	GTG	1,565,633	1,611,246	-73	1,565,846	1,565,773	-73	1,565,527	1,565,454	-73	1,566,723	1,566,650	-73	1,566,186	1,566,113	-73
Gln	I	CTG	1,580,316	1,640,755	70	1,580,529	1,580,599	70	1,580,210	1,580,280	70	1,581,406	1,581,476	70	1,580,869	1,580,939	70
Met	I	CAT	1,600,791	1,681,562	-73	1,601,004	1,600,931	-73	1,600,685	1,600,612	-73	1,601,880	1,601,807	-73	1,601,343	1,601,270	-73
Ala	I	TGC	1,604,532	1,689,045	-72	1,604,745	1,604,673	-72	1,604,426	1,604,354	-72	1,605,621	1,605,549	-72	1,605,084	1,605,012	-72
Ile	I	GAT	1,604,622	1,689,224	-73	1,604,835	1,604,762	-73	1,604,516	1,604,443	-73	1,605,711	1,605,638	-73	1,605,174	1,605,101	-73
Leu	I	CAA	1,693,030	1,693,111	81	1,693,227	1,693,308	81	1,692,880	1,692,961	81	1,694,178	1,694,259	81	1,693,604	1,693,685	81
Ala	I	CGC	1,730,560	1,730,488	-72	1,730,756	1,730,684	-72	1,730,408	1,730,336	-72	1,731,707	1,731,635	-72	1,731,133	1,731,061	-72
Arg	I	CCT	1,774,788	1,774,861	73	1,774,986	1,775,059	73	1,774,617	1,774,690	73	1,775,934	1,776,007	73	1,775,342	1,775,415	73
Val	II	GAC	1,937,268	3,874,465	-71	1,937,522	1,937,451	-71	1,937,097	1,937,026	-71	1,938,393	1,938,322	-71	1,937,881	1,937,810	-71
Asp	II	GTC	2,204,082	4,175,291	-73	2,204,383	2,204,310	-73	2,203,894	2,203,821	-73	2,205,280	2,205,207	-73	2,204,762	2,204,689	-73
Ser	II	TGA	2,438,821	4,416,117	-86	2,439,125	2,439,039	-86	2,438,628	2,438,542	-86	2,440,032	2,439,946	-86	2,439,513	2,439,427	-86
Lys	II	TTT	2,471,261	4,481,011	-72	2,471,565	2,471,493	-72	2,471,068	2,470,996	-72	2,472,472	2,472,400	-72	2,471,953	2,471,881	-72
Cys	II	GCA	2,503,156	4,443,326	70	2,503,459	2,503,529	70	2,502,961	2,503,031	70	2,504,369	2,504,439	70	2,503,853	2,503,923	70
Asn	II	GTT	2,507,591	4,452,055	-71	2,507,895	2,507,824	-71	2,507,396	2,507,325	-71	2,508,804	2,508,733	-71	2,508,288	2,508,217	-71
Thr	II	CGT	2,703,358	4,644,732	72	2,703,647	2,703,719	72	2,703,154	2,703,226	72	2,705,438	2,705,510	72	2,696,937	2,705,009	72
Gly	II	GCC	2,763,467	4,764,949	71	2,763,756	2,763,827	71	2,763,261	2,763,332	71	2,765,547	2,765,618	71	2,765,046	2,765,117	71
Gly	II	GCC	2,893,279	4,821,339	71	2,893,532	2,893,603	71	2,893,073	2,893,144	71	2,895,401	2,895,472	71	2,894,876	2,894,947	71
Leu	II	CAG	2,923,319	4,881,431	83	2,923,578	2,923,661	83	2,923,111	2,923,194	83	2,925,440	2,925,523	83	2,924,916	2,924,999	83
Ser	II	GGA	2,940,747	4,916,290	86	2,941,007	2,941,093	86	2,940,540	2,940,626	86	2,942,868	2,942,954	86	2,942,344	2,942,430	86
Met	II	CAT	2,989,766	4,916,786	-73	2,990,027	2,989,954	-73	2,989,549	2,989,476	-73	2,991,887	2,991,814	-73	2,991,362	2,991,289	-73
Ala	II	TGC	2,993,507	4,924,269	-72	2,993,768	2,993,696	-72	2,993,290	2,993,218	-72	2,995,628	2,995,556	-72	2,995,103	2,995,031	-72
Ile	II	GAT	2,993,597	4,924,448	-73	2,993,858	2,993,785	-73	2,993,380	2,993,307	-73	2,995,718	2,995,645	-73	2,995,193	2,995,120	-73
Glu	II	CTC	3,008,545	4,954,488	71	3,008,806	3,008,877	71	3,008,328	3,008,399	71	3,010,666	3,010,737	71	3,010,141	3,010,212	71
Ser	II	CGA	3,186,829	5,150,177	86	3,187,090	3,187,176	86	3,186,609	3,186,695	86	3,188,945	3,189,031	86	3,188,424	3,188,510	86

tRNA Type	Chr.	Anti-Codon	PT09143			PT09172			Bs143CITA			Bs364CITA			Bs396CITA		
			Begin	End	size	Begin	End	size	Begin	End	size	Begin	End	size	Begin	End	size
Met	II	CAT	3,194,956	5,166,272	-73	3,195,217	3,195,144	-73	3,194,736	3,194,663	-73	3,197,072	3,196,999	-73	3,196,551	3,196,478	-73
Ala	II	TGC	3,198,697	5,173,755	-72	3,198,959	3,198,887	-72	3,198,477	3,198,405	-72	3,200,813	3,200,741	-72	3,200,292	3,200,220	-72
Ile	II	GAT	3,198,787	5,173,934	-73	3,199,049	3,198,976	-73	3,198,567	3,198,494	-73	3,200,903	3,200,830	-73	3,200,382	3,200,309	-73

Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartate; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; Val, valine.

A, adenine; C, cytosine; G, guanine; T, thymine.

Subchapter 4.2.

Evolution and genomic specialization of Brucella suis biovar 2 Iberian Lineages

The results presented in this subchapter were submitted for publication in *BMC Genomics*.

Ferreira, A.C., Tenreiro, R., Corrêa de Sá, M.I. & Dias, R. Evolution and genome specialization of *Brucella suis* biovar 2 Iberian Lineages.

Authors contributions:

ACF and **RD** carried out the genomic comparative analysis and wrote the manuscript.

ACF carried out the experimental work.

MICS, **RT** and **RD** conceived and planned the study, and revised the manuscript.

1. Introduction

Brucella suis is a group of facultative intracellular pathogens that infect a broad range of animals and humans. This species is divided into five biovars, denominated from biovar 1 to 5, that are associated with certain hosts. Considerable diversification within the *B. suis* clade has been observed by different molecular approaches and probable relationships among the biovars and their preferential hosts were suggested (Wattan *et al.*, 2009; Foster *et al.*, 2009; Whatmore *et al.*, 2007; Le Flèche *et al.*, 2006). Biovars 1, 2 and 3 infect suidae and are the etiological agent of swine brucellosis, but the infection caused by biovars 1 and 3 differs from that caused by biovar 2 in host specificity and geographical distribution. In the context of public health, biovar 2 is very rarely pathogenic for humans, whereas biovars 1 and 3 are highly pathogenic causing severe disease in human beings (EFSA, 2009). *B. suis* biovar 2 is found only in Europe and also infects hares, being accept that both the wild boar (*Sus scrofa*) and the European hare (*Lepus europaeus*) are the wild reservoirs of this biovar and the source of transmission of infection to outdoor or extensively reared pigs (Godfroid *et al.*, 2013). Previous studies on the genetic structure of *B. suis* biovar 2 population by Multiple *Locus* Variable-number Tandem Repeat Analysis (MLVA) (Duvnjak *et al.*, 2015; Kreizinger *et al.*, 2014; Li *et al.*, 2013; Garcia-Yoldi *et al.*, 2007; Le Flèche *et al.*, 2006), PCR-RFLP analysis of *omp31*, *omp2a* and *omp2b* genes (Ferreira *et al.*, 2016; Munoz *et al.*, 2010; Garcia-Yoldi *et al.*, 2007; Ferrão-Beck *et al.*, 2006) and whole-genome optical mapping (Ferreira *et al.*, 2016) supported evidence for the existence of two circulating lineages in Iberian Peninsula (Portugal and Spain): the Central-European and the Iberian clonal lineages. The later had been described exclusively in Iberian Peninsula, in both isolates from pigs and wild boars, mainly below the Ebro River suggesting a genomic specialization and local adaptation.

The fitness of a pathogen may vary from host to host. The way that pathogen evolves, within or outside a host, and the strategy used by the host to resist infection can potentially be a determinant for pathogen specialization and an important driver of the evolution of virulence. Host specificity is associated with a number of genomic signatures, including genomic degeneration and genomic rearrangements (Bäumler *et al.*, 2013). Recently, the complete genome sequences of five *B. suis* biovar 2 strains representative of Iberian and Central-European clones, isolated from wild boars in Portugal and Spain, have been released (Ferreira *et al.*, 2014a, 2014b), and it was shown that the Iberian lineage is further characterized by the presence of a large chromosomal inversion (Ferreira *et al.*, 2016).

In order to a better understanding of the mechanisms for evolution and specialization of Iberian lineages, in this study, comparative analysis of *B. suis* biovar 2 with other *Brucella* species were performed to disclose the genomic and structural differences between Iberian and Central-European clones and further discuss the potential factors that favor evolution towards host specialization.

2. Material and Methods

2.1. Bacterial strains and genetic characterization

A total of 190 *Brucella* spp. strains were used for PCR validation assays, including 22 reference strains representative of eight *Brucella* species and 168 *B. suis* isolates, comprising 11 from biovar 1, 152 biovar 2, one biovar 3 and four biovar 4. Among *B. suis* biovar 2 isolates, 104 were representative of the Iberian clonal lineage and 48 of Central-European clonal lineage. Phenotypic characterization of all *Brucella* spp. isolates were performed as previously described (Alton *et al.*, 1988). All *B. suis* isolates were previously subjected to PCR-RFLP analysis of *omp2a* and *omp2b* (Cloeckaert *et al.*, 1995) and *omp31* (Vizcaíno *et al.* 1997) genes, to assess the different haplotypes, and by target-PCR to confirm the presence of the large inversion in *B. suis* biovar 2 Iberian clonal lineage (Ferreira *et al.*, 2016). All data regarding strain characterization can be seen in **Supplementary Tables S2.1 and S2.2 in Subchapter 2.2**, and in **Supplementary Table S3.2** in Chapter 3.

2.2. Comparative genomic analysis

The analysis involved the full sequences of the five newly sequenced genomes of *B. suis* biovar 2 (PT09143, PT09172, Bs143CITA, Bs364CITA and Bs396CITA. For details on sequencing, assembly and annotation see subchapter 4.1), and 18 *Brucella* spp. complete genomes available at the time of analysis from eight of the 12 recognized species: five *B. suis* genomes from biovar 1, 2, 3 and 5; four *B. melitensis* genomes from biovars 1 and 2; three *B. abortus* genomes from biovar 1, including the vaccine strain S19; two genomes from *B. canis* and one each from *B. ovis*, *B. microti*, *B. pinnipedialis* and *B. ceti*. The genome of *Ochrobactrum anthropi* was used as outgroup (**Table 4.2.1**). All DNA sequences were obtained by FTP from the NCBI genome database. Whole-Genome Multiple Sequence Alignment (WG-MSA) of concatenated chromosomes I and II was performed by superstretch approach using as reference the annotated sequence of strain *B. suis* ATCC 23445: DNA seed 10 matches in windows size of 25 bases, minimal stretch length 60 bases, minimal cut-off for stretch identity of 60% in screening windows of 30 bases was used. Clustering analysis was performed using Unweighted Pair Group Method using Arithmetic averages (UPGMA). To allow for accurate alignment between superstretches and avoid false positive mutations, optimized alignment overlapping stretches settings were used. A minimum of 75% of DNA identity and a stretch window of 100 bp were used for determination of orthology. Chromosomal alignment and clustering analysis were conducted in Kodon V3.62 (Applied Maths, Belgium). In order to identify large inserts associated to Iberian *vs* Central-European strains, multiple whole-genome alignment of the sequenced PT09143 plus Bs396CITA strain were made by nucmer3.1 (Kurtz *et al.*, 2004), using as reference strain ATCC 23445.

Single Nucleotide Polymorphism (SNP; including silent, missense and intergenic) and INDEL (insertions and deletions) calling was performed only when superstretch was shown between, at least, the template and query sequence. A window size for scoring mutation clustering of 10 bp was used for mutation quality score. Clustering, distance and score was calculated to assess the significance or quality for each mutation. The number of mutations was counted within the window size. Quality cut-off value was optimized for each alignment by using the 90th percentile of the Poison distribution for score values obtained for each type of mutation. The dN/dS ratio (*i.e.* the ratio of the number of non-synonymous substitutions *per* non-synonymous site to the number of synonymous substitutions *per* synonymous site) was used as an indicator of selective pressure acting on a protein-coding gene.

2.3. Evolutionary whole-genome-based studies

The relationships among *Brucella* spp. genomes were determined using WG-MSA and clustering analysis was performed using UPGMA. All positions containing gaps and missing data were eliminated. Chromosomal alignment and clustering analysis were conducted in Kodon V3.62.

The evolutionary history was inferred from genome-wide SNPs using Neighbour-Joining (NJ) and Maximum Likelihood (ML) methods available in MEGA6 (Tamura *et al.*, 2013). The initial trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. All positions containing gaps and missing data were eliminated. Final phylogenetic trees were obtained using ML method based on the Tamura-Nei model.

The evolutionary history was also inferred from INDELs information obtained from WG-MSA and coded as binary characters (1 if gap present, 0 if absent). Each instance in the 0/1 matrix corresponds to a single INDEL character, which may reflect either an insertion or a deletion relatively to reference sequence. The quality score was optimized based in a set of parameters to discriminate effective INDELS: quality score, INDEL size, number of neighboring SNPs and INDELs, position of the second different base pair, and INDELs distance to the edge on assembled contig. Likewise, only INDELs with quality scores < 1 and sizes > 15 bp were considered. A minimum spanning tree (MST) was generated from a subset of randomize INDELs using BioNumerics version 6.6 (Applied Maths, Belgium).

Table 4.2.1. List of genomes used for comparative genomic and phylogenetic analysis.

NCBI Reference sequence (Chr I/ Chr II)	Strain Identification	Biovar	Host
NC_010169.1/ NC_010167.1	*B. suis* ATCC 23445	2	Hare
NZ_CP007691.1/ NZ_CP007692.1	*B. suis* PT09143	2	Wild boar
NZ_CP007693.1/ NZ_CP007694.1	*B. suis* PT09172	2	Wild boar
NZ_CP007695.1/ NZ_CP007696.1	*B. suis* Bs143CITA	2	Wild boar
NZ_CP007697.1/ NZ_CP007698.1	*B. suis* Bs364CITA	2	Wild boar
NZ_CP007720.1/ NZ_CP007721.1	*B. suis* Bs396CITA	2	Wild boar
NC_016797.1/ NC_016775.1	*B. suis* VBI22	1	Swine
NC_004310.3/ NC_004311.2	*B. suis* 1330	1	Swine
NC_015857.1/ NC_015858.1	*B. pinnipedialis* B2/94	Not applicable	Dolphin
NZ_CP007719.1/ NZ_CP007718.1	*B. suis* bv.3 str. 686	3	Seal
NC_016778.1/ NC_016796.1	*B. canis* HSKA52141	Not applicable	Dog
NC_010103.1/ NC_010104.1	*B. canis* ATCC 23365	Not applicable	Dog
NC_013119.1/ NC_013118.1	*B. microti* CCM 4915	Not applicable	Wild rodent
NZ_CP007717.1/ NZ_CP007716.1	*B. suis* 513UK	5	Wild boar
NC_022905.1/ NC_022906.1	*B. ceti* TE10759-12	Not applicable	Seal
NC_007618.1/ NC_007624.1	*B. abortus* 2308	1	Cattle
NC_006932.1/ NC_006933.1	*B. abortus* 9-941	1	Cattle
NC_010742.1/ NC_010740.1	*B. abortus* S19	1	Vaccine
NC_012441.1/ NC_012442.1	*B. melitensis* ATCC 23457	2	Goat
NC_009505.1/ NC_009504.1	*B. ovis* ATCC 25840	Not applicable	Sheep
NC_017246.1/ NC_017246.1	*B. melitensis* M5-90	1	Sheep
NC_017244.1/ NC_017244.1	*B. melitensis* M28	1	Sheep
NC_003317.1/ NC_003318.1	*B. melitensis* 16M	1	Goat
NC_009667.1/ NC_009668.1	*Ochrobactrum anthropi* ATCC 49188	Not applicable	Not applicable

2.4. PCR assessment of INDEL events

All the INDELs found that were discriminative of *B. suis* biovar 2 Iberian clonal lineage were assessed by PCR in the collection of the 190 *Brucella* strains previously described. Primers for each lNDEL were designed using the nucleotide positions relative to the reference strain ATCC 23445 and specificity of all the primers was confirmed by BLAST analysis against the published genome sequence of the reference strain. The primer sequences and expected amplicon sizes are provided in **Table 4.2.2**. Genomic DNA from each strain was prepared using the High Pure PCR Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) according to the manufacturer. PCR amplification was performed in a total volume of 25 µl containing 25 ng of DNA, 1X PCR reaction buffer, 1 U of Taq DNA polymerase (Promega, USA), 200 mM of each dNTPs and 0.3 mM of each flanking primers. Amplifications were performed in a

MyCycler thermal Cycler (Bio-Rad, France). An initial denaturation step at 95 ºC for 3min was followed by 30 cycles of denaturation at 95 ºC for 20 s, primer annealing at 56 ºC for 30 s and elongation at 72 ºC for 30 s. The final extension step was performed at 72 ºC for 5 min. Five microliters of amplification products were loaded on a 2% standard agarose gel and run under a voltage of 8 V/cm for 60-90 min. A 100 bp ladder (Invitrogen, USA) was used as molecular size marker.

4.2.2. List of the primers used for assessment of the INDEL events differentiating the two *B. suis* biovar 2 clonal lineages.

CE, Central-European strains; IB, Iberian strains.

INDEL	Primer sequence 5' → 3'	Amplified region in ATCC 23445	Chr	Affected CDS in ATCC 23445	PCR product expected size (bp)	
					CE	IB (only 2e)
SI79506	F tgcacacgtagggtcgata R cacccagatgttcggctat	79,420..80,357	I	BSUIS_A0075	938	260
SI1356057	F taggctgcctggaattcatc R cggcaagttcacctctgact	1,355,436..1,356,288	I	Intergenic	853	765
SI1423448	F tttcttattccacccgatcc R caaaatgttactgcgtgaagc	1,423,358..1,423,655	I	Intergenic	298	298 (218)
SI1627485	F caatccacaggagatcggt R gcttcacgggtatcatgtaca	1,627,421..1,627,600	I	BSUIS_A1714	180	129
SI2041144	F tcggaaatggacgaatatca R ttcttgtcgtcggaaatgtc	107,306..107,533	II	Intergenic	228	228 (171)
SI2603410	F agtggattttggtgcgtttc R aagatgagcgggaaatgttg	679,423..680,774	II	BSUIS_B0700; BSUIS_B0701	1,352	508
LI3234619	F tattattcactttgagcggca R aactgcaaaagcttggctg	1,310,801..1,312,970	II	BSUIS_B1354-BSUIS_B1357	2,170	174

3. Results and Discussion

3.1. Phylogenomic relationships of *B. suis* biovar 2

To understand the genomic specialization observed in Iberian Peninsula, the genetic structure and evolutionary relationships of the *B. suis* strains isolated at Iberian Peninsula were assessed by three Whole-Genome based phylogeny approaches: Multiple Sequence Alignment (WG-MSA); Single Nucleotide Polymorphism distribution (WG-SNP) and Insertion and deletion distribution (WG-INDEL). The analysis involved the full sequence of the five new genomes and 18 *Brucella* complete genomes. All genomes with the exception of *B. suis* 686 (biovar 3) have two circular chromosomes (**Table 4.2.1**).

The WG-MSA analysis was performed for the concatenated chromosomes of each genome and the resultant similarity matrix was used for clustering of the 23 genomes (**Figure 4.2.1**). The UPGMA dendrogram grouped *Brucella* spp. genomes into two highly distinct clusters (I and II) according to their nucleotide sequence identity. Cluster I grouped the genomes of three *B. abortus* strains, *B. suis* 686 (biovar 3) and three strains of *B. suis* biovar 2 Iberian clonal lineage (PT09172, PT09143 and Bs143CITA). Cluster II gathered the remaining 16 *Brucella* genomes. The average nucleotide identity (ANI) between cluster I and cluster II was 74%, which was lower than expected. ANI is one of the most robust measurements of genomic relatedness between strains. However, genomic structural organizations such as chromosomal inversions or translocations events may affect the identity values even though little impact is observed in gene content (Chun & Rainey, 2014). Indeed, *B. abortus* presents an inversion of 640 kb in chr II (Foster *et al.,* 2009; Halling *et al.*, 2005; Michaux-Charachon *et al.*, 1997) and *B. suis* biovar 2 strains PT09172, PT09143 and Bs143CITA present an inversion in chr I with approximately 944 kb (Ferreira *et al.*, 2016), as well as the known 210-kb segment of chr I translocated to chr II (Ferreira *et al.*, 2016; Foster *et al.*, 2009; Wattan *et al.*, 2009). Moreover, strain 686 from *B. suis* biovar 3, the unique reference strain with a single chromosome of 3.3 Mbp (Jumas-Bilak *et al.*, 1998), also revealed an inversion of 790 kb, located between nucleotide #285,192 and #1,074,164 (ATCC 23445 reference positions), in a similar region as the inversion occurring in biovar 2 strains from the Iberian clonal lineage (cluster I). To the best of our knowledge, this is the first description of a genomic inversion in *B. suis* biovar 3 and of chromosomal structural rearrangements as evolutive features shared by biovar 2 and biovar 3 strains.

To unveil the evolutionary history of *B. suis* biovar 2 imprinted on the genomic background, WG-SNP analysis was performed to remove the effect of chromosomal rearrangements (translocations and inversions) on *B. suis* phylogeny. Whole-genome alignment of the 23 *Brucella* strains yielded a total 30,255 and 332,119 putative SNPs in comparison to the reference genome and outgroup strain, respectively (data can be found in **Supplementary Table S4.2.1-A**, allocated at http://www.mediafire.com/file/2mm9zdwo7lyovm2/Supplementary_Table_S4.2.1_A-C.xlsx). The evolutionary distances were computed assuming equality of substitution pattern among lineages and of substitution rates among sites using the Maximum Composite Likelihood method and clustered by Neighbor-Joining method (**Figure 4.2.2**). The unrooted tree sorts the *Brucella* genomes into eight clades, as follows: the *B. suis* biovar 2 clade (A); the *B. suis-B. canis* clade (B); the *B. suis* biovar 5 clade (C); the *B. microti* clade (D); the *B. ovis* clade (E); the *B. pinnipedialis-B. ceti* clade (F); the *B. abortus* clade (G), and the *B. melitensis* clade (H) (**Figure 4.2.2**). These results are in accordance with several phylogenetics studies using WG-SNP or MLST data (Sankarasubramanian *et al.*, 2016; Foster *et al.,* 2012; Audic *et al.*, 2009; Wattan *et al.*, 2009; Foster *et al.*, 2009; Whatmore *et al.*, 2007; Chain *et al.*, 2005). The *B. suis*

biovar 2 isolates form two distinct phylogenetic groups in clade A, one corresponding to the Central-European clonal lineage (subclade A1) and the other to the Iberian clonal lineage (subclade A2), thus giving additional phylogenetic support at genome level for the occurrence of these clonal lineages already revealed by chromosome I optical maps (Ferreira *et al.*, 2016). Moreover, the evolutionary history inferred from the 944 kb inversion region distinctive of subclade A2 and from 210 kb translocated region of clade A showed no differences in comparison to whole-genome analysis. Alike topologies were observed for coding and non-coding regions (data not shown). Thus, the mechanisms for genomic specialization on the Iberian lineages seem to be independent of the chromosomal rearrangement events observed in these clades.

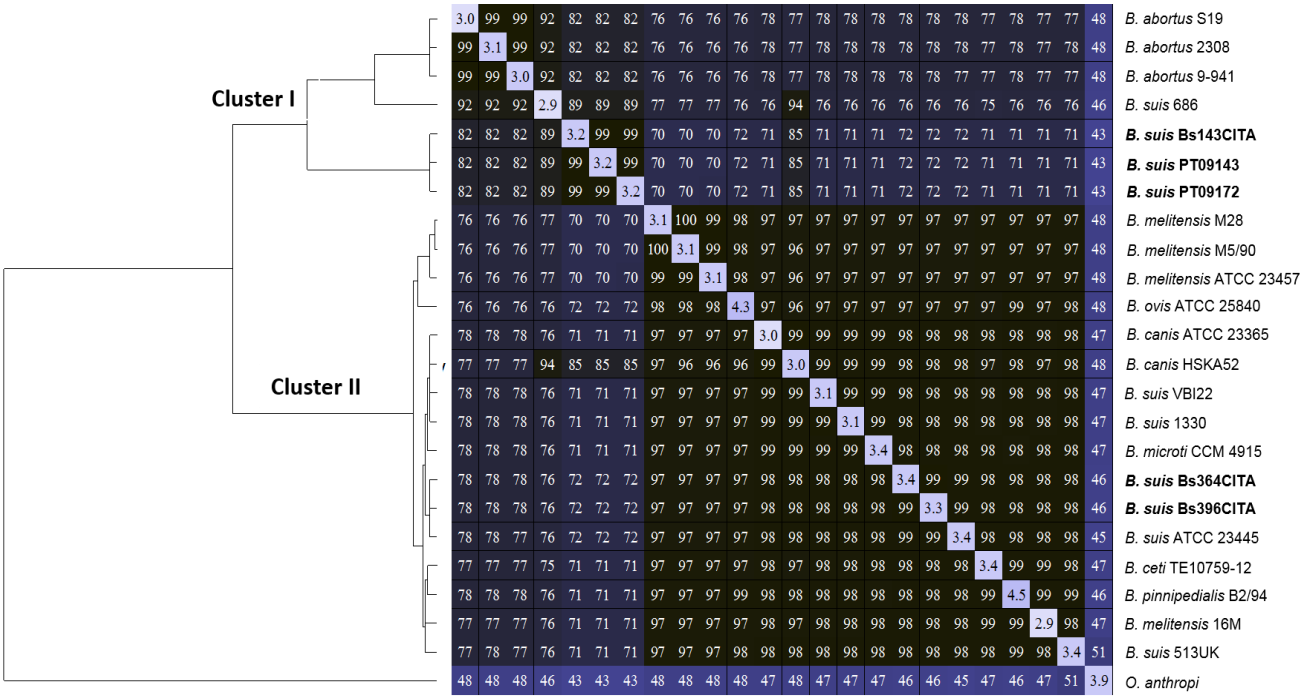


Figure 4.2.1. Comparative chromosome mapping of 23 *Brucella* spp. genomes. Genomic alignment of concatenated chromosomes I and II was performed by superstretch approach: DNA seed 10 matches in windows size of 25 bases, minimal stretch length 60 bases, minimal cut-off for stretch identity of 60% in screening windows of 30 bases was used. Each cell in the matrix displays the identity score, with a corresponding color scale. The scale goes from black, corresponding with 100% identity, over blue towards white (0% identity). The left-to-right diagonal of the matrix contains those cells representing the comparison of sequences compared to themselves. The value in each cell represent the percentage of repetitive regions for that sequence. Clustering analysis was performed using UPGMA. All positions containing gaps and missing data were eliminated. The genome of *O. anthropi* was used as outgroup.

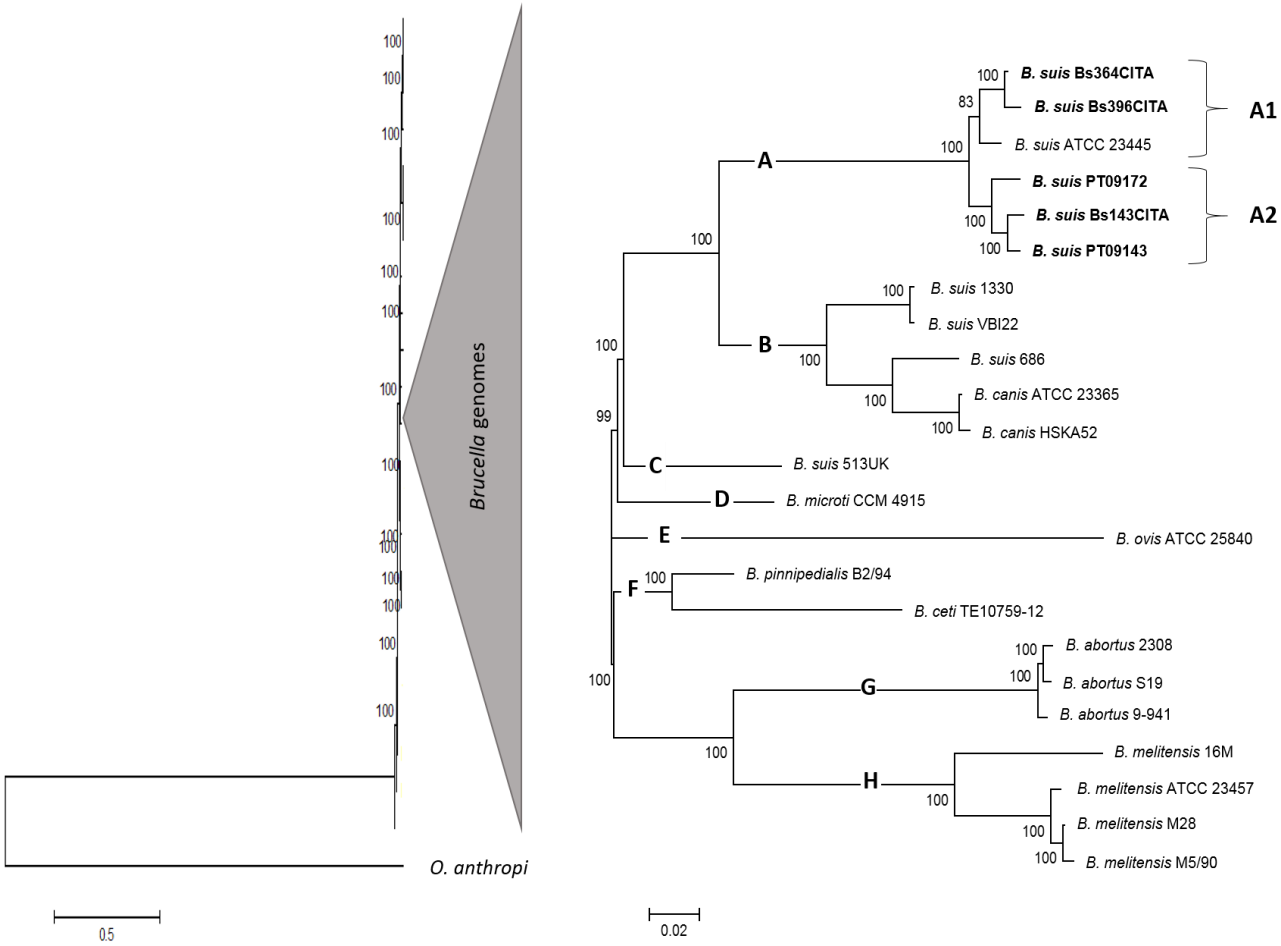


Figure 4.2.2. Evolutionary relationships inside genus *Brucella* inferred from WG-SNP analysis. Rooted phylogenetic tree with *O. anthropi* as outgroup; all positions containing gaps and missing data were eliminated and a total of 332,119 positions was used in the final dataset. The unrooted *Brucella* phylogenetic tree is shown in more detail, involving a total of 30,255 positions in the final dataset and depicting eight clades (A to H). The evolutionary history was inferred by using the Maximum Likelihood method. The tree with the highest log likelihood (-193253.4180) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Joining and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions *per* site. Evolutionary analyses were conducted in MEGA6.

The WG-SNP approach relies on reference-based mapping and do not detect variations in regions that are not present in the reference sequence (*i.e.* ATCC 23445), being suitable only for comparison of the core genome shared among all strains. WG-INDELs are largely ignored in phylogenetic reconstruction but provide a suite of markers complementary to nucleotide substitutions with enormous potential for molecular phylogenetics (Ashkenazy *et al.*, 2014; Luan *et al.*, 2013; Redelings and Suchard, 2007). In order to confirm the phylogenetic relationships suggested by the former

approaches, a minimal spanning tree (MST) was generated from the 3,052 effective INDELs (data can be found in **Supplementary Table S4.2.1-B,** http://www.mediafire.com/file/2mm9zdwo7lyovm2/Supplementary_Table_S4.2.1_A-C.xlsx).

A total of 1,156 INDELs were unique to *O. anthropi* and 1,896 INDELs were found among *Brucella* species. The MST was built using a data set of 255 representative INDELs, which included seven Iberian specific INDELs and a subset of 248 randomly chosen INDELs from the 3,052. Evolutionary relationships based on WG-INDELs are congruent with those from WG-SNP analysis, further revealing *B. suis* biovar 2 Central-European clonal lineage genomes allocated in a well-defined cluster, from which the other *B.suis* biovars and *Brucella* species seems to evolve (**Figure 4.2.3**). Therefore, we can speculate that biovar 2 Iberian clonal lineage evolved from the Central-European clonal lineage, representing an on-going allopatric speciation process as described for other specialized pathogenic bacteria (Georgiades & Raoult, 2010).

Brucella species are able to infect multiple hosts but *B. suis* biovar 2 is highly adapted to wild boars (Olsen & Palmer, 2014). In fact, the results obtained from the three whole-genome approaches suggest the genomic specialization of *B. suis* biovar 2 on the Iberian Peninsula to be independent of a specific genomic event(s) but instead driven by host specialization, establishing an ecovar.

3.2. Comparative genomics of *B. suis* biovar 2

High homology (>98%) between *Brucella* spp. are normally found in association with their preferred hosts and has apparently resulted in adaptive changes over time (Ficht, 2010). Phylogenomic relationships have supported the existence of two *B. suis* biovar 2 phylogenetic groups well separated from *B. suis* biovar 1 and 3. The comparative genomics analysis was focused on the detection of distinctive genetic events between those two lineages. The inter- and intraspecies comparative analysis revealed several biovar-, haplotype- and strain-specific genetic polymorphisms that can implicate further genetic determinants related to host specificity and genomic specialization on *B. suis*. A set of chromosomal rearrangements and polymorphisms, including SNPs and INDELs, were found and will be further discussed in the following sections.

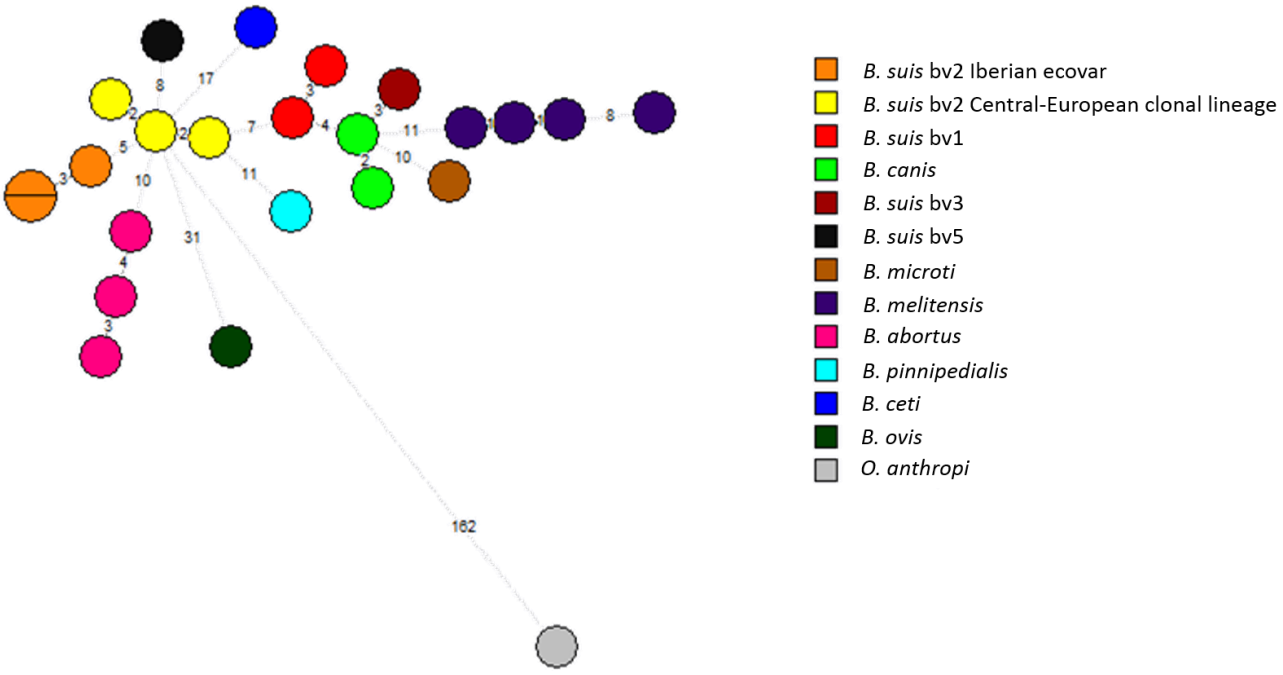


Figure 4.2.3. Minimum spanning tree depicting the genomic relationships of *Brucella* species based on WG-INDELs analysis. A data set of 255 representative INDELs was used, including 7 Iberian ecovar specific INDELs and a subset of 248 randomly chosen INDELs from the 3,052 INDELs obtained from the comparative analysis of 23 *Brucella* genomes and *O. anthropi* (outgroup). Colour codes are associated with *Brucella* spp. groups and the number of changes between patterns is presented. The MST was constructed with a categorical coefficient using BioNumerics version 6.6.

3.2.1. Large chromosomal rearrangements in *B. suis* biovar 2 genomes

The chromosomal organization of the six *B. suis* biovar 2 strains was examined by *BamH* I optical mapping and the pairwise alignment between optical and *in silico* maps allowed the identification of two major chromosomal rearrangements (one translocation event and one inversion) occurring in *B. suis* biovar 2 genomes, pointing out the degree of genome plasticity in *B. suis* species (Ferreira *et al.*, 2016).

The 210 kb region translocated from Chr I to Chr II, associated to *IS*711, in the six biovar 2 studied genomes (**Figure 4.2.4**) was firstly described in *B. suis* biovar 2 and 4 reference strains (Jumas-Bilak *et al.*, 1998). The translocate region share 99% similarity with *B. suis* 1330 and encompasses genes with most functions assigned to processes such as transcription, replication and repair, carbohydrate metabolism and metabolism of co-factors and vitamins. Moreover, five tRNAs genes and six ribosomal proteins (L21, L27, L31, L32, L36 and S16) were moved to Chr II, and 14 coding regions in *B. suis* biovar 2 don't have orthologs in *B. suis* 1330 genome.

The most surprising rearrangement found was the 944 kb chromosomal inversion present in strains PT09143, PT09172 and Bs143CITA (*B. suis* biovar 2 clade, subcluster A1, Iberian clonal lineage) and

covering 49% of the Chr I (Ferreira *et al.*, 2016). Chr I encodes the majority of the core metabolic machinery for processes such as transcription, translation, and protein synthesis (Paulsen *et al.*, 2002). Although changing gene location, the majority (>95%) of annotated coding regions were found to share 98-100% sequence identity with ATCC 23445 and both genomes from the Central-European clonal lineage.

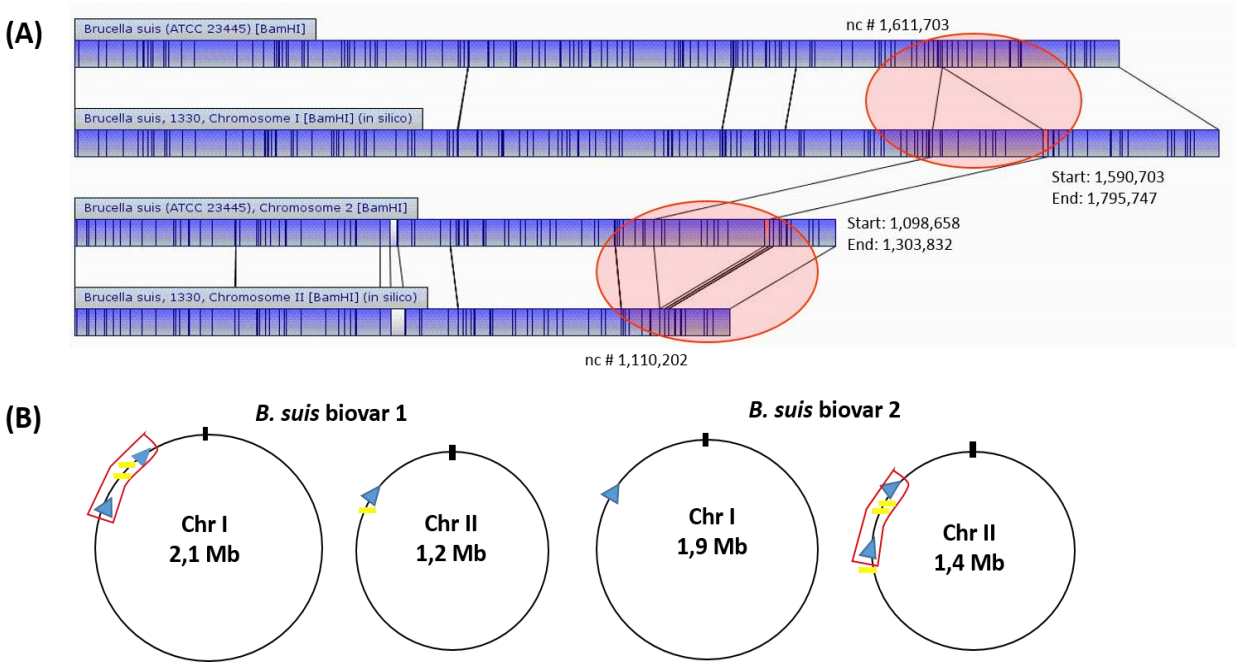


Figure 4.2.4. Translocation event in *B. suis* biovar 2 strains. **(A)** Pairwise alignment of chromosome I and II optical and *in silico* maps for *B. suis* biovar 2 strain ATCC 23445 and *B. suis* biovar 1 strain 1330. Lines connecting two chromosomal maps indicate discontinuity in the alignment of fragments. The translocated region is highlighted in the red circles; unaligned restriction fragments, representing differences between two chromosomes, are showed in white; blue indicates aligned restriction fragments. **(B)** Schematic representation of both circular chromosomes. Open red box indicates the translocated region in chromosome I of *B. suis* biovar 1 and in chromosome II of *B. suis* biovar 2; blue triangles symbolize the *rrn loci*; yellow boxes represents the insertion sequence *IS*711.

At left and right crossover points, the inversion disrupted a Tripartite ATP-independent periplasmic (TRAP) dicarboxylate transporter, DctM subunit (C-terminus truncated, ortholog to ATCC 23445 BSUIS_A0375 and 1330 BR0344; nucleotide #371,383 in ATCC 23445), and an integral membrane protein TerC (N-terminus truncated, ortholog to ATCC 23445 BSUIS_A1382 and 1330 BR1332; nucleotide #1,316,165 in ATCC 23445), respectively (**Figure 4.2.5**). The TRAP transporters are a large family of substrate-binding protein (SBP)-dependent secondary transports found in bacteria and archaea. These transporters have three domains that were defined on the basis of orthology to the three proteins that constitute the Dct system in *Rhodobacter capsulatus* (Forward *et al.*, 1997), comprising a SBP of the

DctP or TAXI families and two integral membrane proteins that form the DctQ and DctM protein families. Orthologs of the three genes can be found in all *Brucella* species (BSUIS_A0374 to BSUIS_A0376 orthologs in ATCC 23445) although *dctQ* and *dctP* were found to be variable within *B. abortus*, *B. canis*, and *B. melitensis*. The gene *dctM* is well conserved in *Brucella* spp. except in *B. suis* biovar 2 strains from Iberian ecovar, probably resulting in the inactivation of this gene in those strains. In *B. melitensis* 16M the Dct operon (BMEI1579- BMEI11581) was predicted to be involved in the transport of mannitol although no experimental evidences exist, but in *R. capsulatus*, *dctP, dctQ,* and *dctM* genes were shown to be essential for C4-dicarboxylate transport (Forward *et al.*, 1997). Mannose is both an important precursor in the O-antigen biosynthetic pathway and in the production of the inner core moiety of lipopolysaccharide (LPS) (Cardozo *et al.*, 2006; Zygmunt *et al.*, 1988). Loss of the ability to uptake mannitol can influence LPS structure and subsequently host immune responses. At the right crossover, the interrupted ORF codes for a TerC, a protein possibly involved in tellurium resistance (inorganic ion transport and metabolism). This membrane protein harbors a CBS domain that is usually associated to enzymatic domains, membrane transporters or DNA-binding domains, playing an important role in host interactions.

Taking into consideration the genomic rearrangements occurring in biovar 2 genomes and the possible effects in whole genome structure, it can be accepted that those events were tolerated from an evolutionary point of view, but probably changed the metabolism of *B. suis* biovar 2 strains resulting in host adaptation. In fact, chromosomal divergence has been detected in related species, and this might create bacteria sufficiently different to escape the immune system and establish an infection in new hosts (Hughes, 2000).

3.2.2. Distinctive SNPs of *B. suis* biovar 2 Iberian ecovar and functional assessment

The number of SNPs, including intergenic, missense and silent, obtained amongst the *Brucella* genomes relatively to *B. suis* ATCC 23445 are shown in **Table 4.2.3**. Eighty-five percent of the SNPs were identified in coding regions and in average 55% are missense mutations, affecting genes associated with different classes of cellular functions. The distribution of SNPs along the genome (SNPs per 200 kb) showed no evident differences for the aforementioned chromosomal inversion at Chr I from Iberian *B. suis* biovar 2 strains (**Supplementary Figure S4.2.1,** http://www.mediafire.com/file/3m9xsgpsfbf17tx/Supp.Figure_S4.2.1_SNPs_distribution.pdf). The five *B. suis* biovar 2 strains shared 4,107 SNPs. Two-hundred and seven SNPs were unique to strains belonging to the Iberian ecovar (PT09143, PT09172 and Bs143CITA) and 232 discriminate the Central-European strains Bs364CITA and Bs396CITA. Among these, 104 missense SNPs were found specifically in strains from Iberian ecovar.

In the three strains from the Iberian ecovar was observed a relative and preferential accumulation of missense SNPs among the annotated genes associated to transporter/ efflux systems, other membrane receptors, metabolism processes, transcriptional regulators, regulatory proteins, cell replication, SOS response/DNA repair and ribosomal proteins (**Figure 4.2.6**). The majority of discriminative mutations were associated to membrane related molecules (29%; n= 30) and enzymes involved in catabolism process (20%; n=21). These mutations can be related to molecular tropism to a specific environment or animal host.

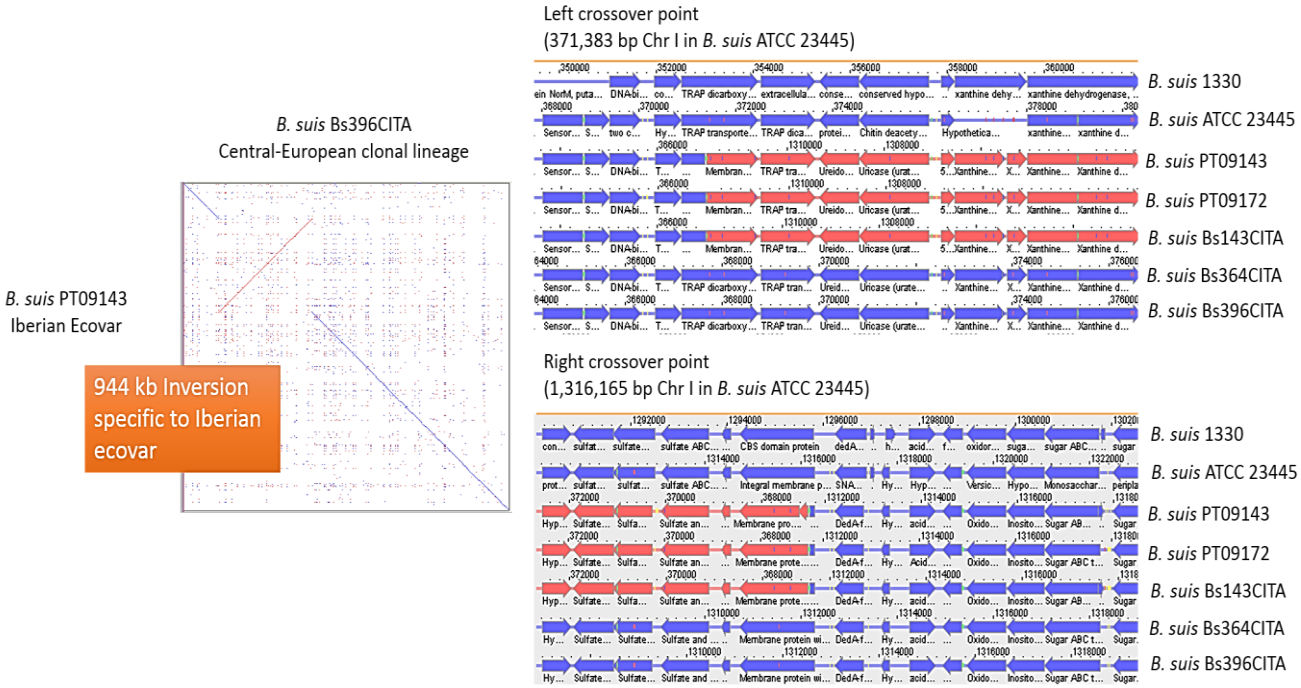


Figure 4.2.5. Left and right crossover points of the large inversion present in *B. suis* biovar 2 Iberian ecovar. Dot-blot graphic demonstrating the presence of the 944 kb inversion in Chr I of *B. suis* biovar 2 strains from Iberian ecovar (1.9 Mb, aprox. 49%). As an example, the comparison of *B. suis* PT09143 (as the representative strain from Iberian ecovar) with *B. suis* Bs396CITA (representative of the Central-European clonal lineage) is shown. The right and left crossover points are shown. The inversion disrupted a Tripartite ATP-independent periplasmic (TRAP) dicarboxylate transporter, DctM subunit (nucleotide #371,383 in ATCC 23445), and an integral membrane protein TerC (nucleotide #1,316,165 in ATCC 23445), respectively. Chromosomal alignment was conducted in Kodon V3.62.

Table 4.2.3. Resume of mutation analysis of *Brucella* spp. genomes using *B. suis* ATCC 23445 as reference.

NCBI Reference sequence (Chr I/ Chr II)	Strain Identification	Biovar	Host	INDELs	SNPs			
					Int.	Mis.	Sil.	Total
NC_010169.1/ NC_010167.1	*B. suis* ATCC 23445	2	H	-	-	-	-	-
NZ_CP007697.1/ NZ_CP007698.1	*B. suis* Bs364CITA	2	WB	59	171	269	148	588
NZ_CP007720.1/ NZ_CP007721.1	*B. suis* Bs396CITA	2	WB	68	271	296	167	734
NZ_CP007693.1/ NZ_CP007694.1	*B. suis* PT09172	2	WB	79	285	398	241	924
NZ_CP007691.1/ NZ_CP007692.1	*B. suis* PT09143	2	WB	77	232	438	263	933
NZ_CP007695.1/ NZ_CP007696.1	*B. suis* Bs143CITA	2	WB	104	307	451	270	1,028
NC_016797.1/ NC_016775.1	*B. suis* VBI22	1	SW	358	719	2,716	1,540	4,975
NC_004310.3/ NC_004311.2	*B. suis* 1330	1	SW	373	737	2,669	1,573	4,979
NC_015857.1/ NC_015858.1	*B. pinnipedialis* B2/94	na	D	470	781	2,962	1,547	5,290
NZ_CP007719.1/ NZ_CP007718.1	*B. suis* bv.3 str. 686	3	SW	395	773	2,986	1,680	5,439
NC_016778.1/ NC_016796.1	*B. canis* HSKA52141	na	Dog	388	808	3,020	1,649	5,477
NC_010103.1/ NC_010104.1	*B. canis* ATCC 23365	na	Dog	390	836	2,989	1,723	5,548
NC_013119.1/ NC_013118.1	*B. microti* CCM 4915	na	WR	384	1,023	2,865	1,676	5,564
NZ_CP007717.1/ NZ_CP007716.1	*B. suis* 513UK	5	WR	400	855	3,067	1,643	5,565
NC_022905.1/ NC_022906.1	*B. ceti* TE10759-12	na	S	576	1,027	3,804	2,053	6,884
NC_007618.1/ NC_007624.1	*B. abortus* 2308	1	C	601	1,172	4,488	2,494	8,154
NC_006932.1/ NC_006933.1	*B. abortus* 9-941	1	C	595	1,172	4,508	2,482	8,162
NC_010742.1/ NC_010740.1	*B. abortus* S19	1	Vac	610	1,159	4,515	2,493	8,167
NC_012441.1/ NC_012442.1	*B. melitensis* ATCC 23457	2	G	581	1,216	4,717	2,614	8,547
NC_009505.1/ NC_009504.1	*B. ovis* ATCC 25840	na	SH	815	1,239	4,569	2,745	8,553
NC_017246.1/ NC_017246.1	*B. melitensis* M5-90	1	SH	635	1,218	4,744	2,613	8,575
NC_017244.1/ NC_017244.1	*B. melitensis* M28	1	SH	589	1,226	4,750	2,650	8,626
NC_003317.1/ NC_003318.1	*B. melitensis* 16M	1	G	675	1,141	5,641	2,031	8,813

na, not applied.

Host: C, cattle; D, dolphin; G, goat; H, hare; S, seal; SH, sheep; SW, swine; WB, wild boar; WR, wild rodent; Vac, vaccine.

SNPs: Int., intergenic; Mis., missense; Sil., silent.

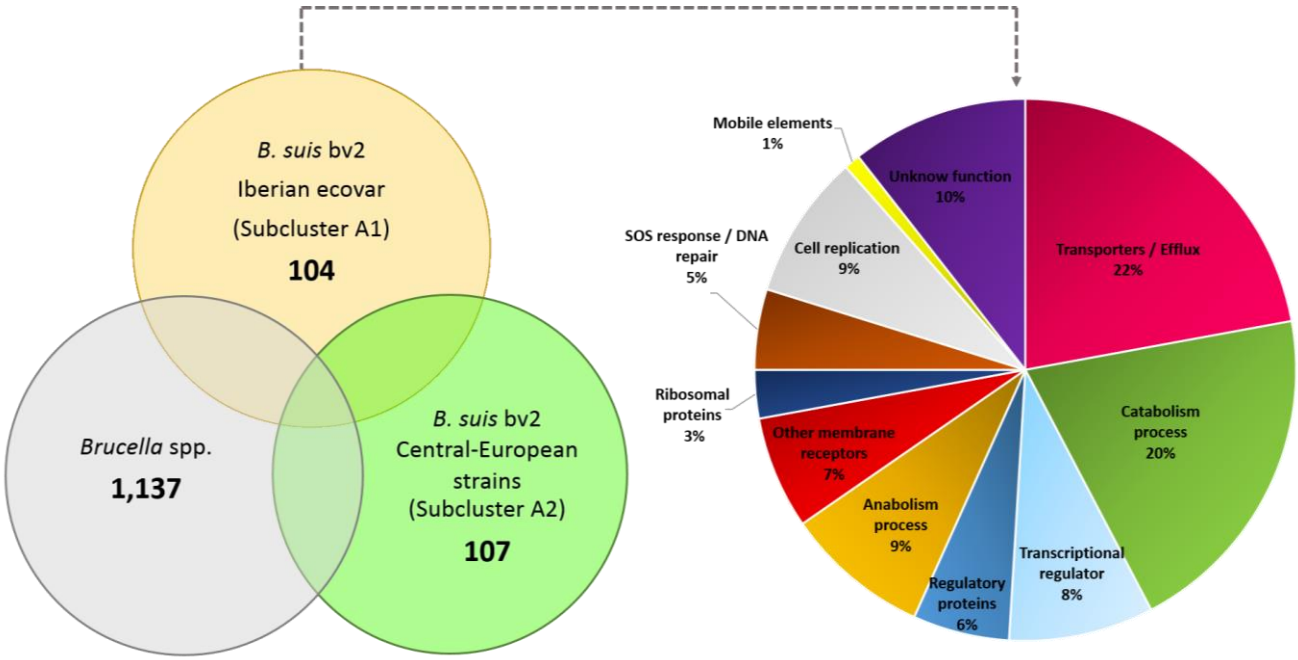


Figure 4.2.6. Distribution of putative missense mutations between the two distinct *B. suis* biovar 2 ecovars. The Venn diagram shows the number of unique SNPs between *B. suis* biovar 2 Iberian ecovar (PT09143, PT09172 and Bs143CITA), *B. suis* biovar 2 Central-European lineage (Bs364CITA and Bs396CITA) and the remaining *Brucella* spp. genomes. The associated pie chart shows the breakdown of the functional categories assigned to unique SNPs of Iberian ecovar.

3.2.3. INDELs differentiative of biovar 2 Iberian ecovar

The comparative analysis of the effective INDELs disclosed 10 INDELs in coding regions, including one Large INDEL (LI ≥ 1 kb), and three intergenic INDELs were found to be specific of strains PT09143, PT09172 and Bs143CITA representative of Iberian ecovar (**Supplementary Table S4.2.1-B,** http://www.mediafire.com/file/2mm9zdwo7lyovm2/Supplementary_Table_S4.2.1_A-C.xlsx). The LI with, 1,996 bp in Chr II (LI3234619), is present in one region encoding ORFs BSUIS_B1354 to BSUIS_B1357 in ATCC 23445 within one *Brucella* flagellar gene cluster. BSUIS_B1354, BSUIS_B1355 and BSUIS_B1357 (BRA1127, BRA1128 and BRA1130 orthologs in 1330) code for hypothetical proteins without any specific functional domains. BSUIS_B1356 (BRA1129 ortholog in 1330) is predicted to code a flagellar protein FlgJ, possessing an N-terminal domain responsible for proper rod assembly. Primers (LI3234619-F and LI3234619-R) directed to flanking regions were used for INDEL search, confirming the presence of this INDEL in strains PT09143, PT09172, Bs143CITA and in the 104 biovar 2 strains from Iberian ecovar. Although brucellae were considered non-motile bacteria for a long time, it was recently proved that *B. melitensis* produces a functional flagellum with the characteristics of a sheathed flagella described in other organisms, produced only transiently at the end of the exponential phase of growth. The bacterial flagellum is a complex apparatus assembled of at least 31 different proteins, with similar

organization in all *Brucella* genomes with genes distributed in three clusters on the small chromosome (Ferooz and Letesson, 2010; Chain *et al.*, 2005; Fretin *et al.*, 2005). It is documented that flagellar genes are required for the establishment of *in vivo* infection in mice and goats (Zygmunt *et al.*, 2006; Fretin *et al.*, 2005). Consequently, the inactivation (due to point mutations or INDELs) or loss of key flagellar genes would influence the formation of a functional flagellum and therefore several functions can be affected, such as protein export or adhesion (Haiko and Westerlund-Wikström *et al.*, 2013).

Among the nine small INDELs (SI<1 kb) in coding regions, three specific SIs were further evaluated due to their location in genome and eventual importance in evolution and genomic specialization of *B. suis* biovar 2 Iberian ecovar. SI79420 (678 bp) occurs within a CDS coding for an outer membrane protein ortholog to ATCC 23445 BSUIS_A0075 and 1330 BR0072, which have been described as a putative autotransporter adhesin (Chain *et al.*, 2005; Paulsen *et al.*, 2002). In fact, different proteins belonging to autotransporter family have been identified in *Brucella* genomes, sharing a common domain organization: an N-terminal secretion signal, a divergent and functional domain (passenger domain) and a conserved C-terminal region (Ruiz-Ranwez *et al.*, 2013a, 2013b; Posadas *et al.*, 2012; Tsolis *et al.*, 2009; Chain *et al.*, 2005). The alignment of the orthologs genes in the different species indicated that this INDEL occurs within the passenger domain, showing a range in similarity (at the nucleotide level) to BSUIS_A0075, from 98% (Bs364CITA and Bs396CITA), to 72% similarity (PT09143, PT09172 and Bs143CITA). SI79420 probably caused the inactivation of the protein but it remains to be seen if this protein is functional or if differences within the passenger domain contribute to host or tissue specificity or clinical manifestations in wild boars or pigs. SI1627421 (59 bp) affects a permease, ortholog to ATCC 23445 BSUIS_A1714 and 1330 BR1873, associated to autotransporter proteins. The three strains from Iberian clonal lineage presented the 59 bp INDEL next to the C-terminus. Since almost all of the known autotransporters are involved in functions related to the invasion process, the difference in the number of active autotransporters, and the variation within them, may play a role in the ability of each species to interact with its host and may thus be an important contributor to virulence (Chain *et al.*, 2005; Paulsen *et al.*, 2002). Lastly, SI2603410 (844 bp) represent an event that causes the elimination of one insertion sequence. The *IS711* insertion sequence is unique to *Brucella* species, and its copy number in the genome varies between species and biovars, being regarded as key determinants of genome plasticity and have been suggested to provide significant adaptive changes to genomes. Seven complete copies of this insertion sequence are recognized in *B. suis* 1330 and 13 in ATCC 23445 (Audic *et al.*, 2011), as well as in strains Bs364CITA and Bs396CITA. However, strains PT09143, PT09172 and Bs143CITA present 12 copies of this insertion element, including *orfA* and *orfB* genes. From the three intergenic INDELs, the SI1356057 (88 bp) is located between a GntR family transcriptional regulator and a ketol-acid reductoisomerase. In ATCC 23445, a CDS coding for a hypothetical protein (BSUIS_A1430) is

annotated among the GntR family transcriptional regulator (BSUIS_A1429), and the ketol-acid reductoisomerase (BSUIS_A1431). Orthologous of the three genes are also present in both Central–European strains, but BSUIS_A1430 ortholog is missing in PT09143, PT0172 and Bs143CITA. Additionally, two intergenic SIs specific for strains PT09143 and Bs143CITA (Iberian ecovar, haplotype 2e), one in Chr I with 77 bp (SI1423448), and other in Chr II with 47 bp (SI2041144) were found. Nevertheless, no frameshift is expected or promoter region seemed to be affected by those two described INDELs.

The six abovementioned SIs were searched by targeted-PCRs in the 190 *Brucella* strains and it was confirmed that those events were specific of Iberian ecovar (**Supplementary Table S4.2.2**).

Lastly, no INDELs were found to affect genes known to participate in virulence, such as lipopolysaccharide biosynthesis, two-component regulatory system BvrR/BvrS, type IV secretion system VirB or erythritol catabolic pathway.

4. Conclusion

In this work, a full genome comparative analysis of five *B. suis* biovar 2 strains isolated from wild boars belonging to the main circulating clonal lineages in Iberian Peninsula together with the publicly available *Brucella* spp. genomes was performed. *Brucella* spp. are genetically high conserved and have apparently slow evolutionary rates. Nevertheless, it was observed that *B. suis* biovar 2 strains from Iberian clonal lineage are differentiated from those from Central-European clonal lineage not only by the presence of one large inversion in Chr I but also by a number of specific SNPs, deletions and insertions. Additionally, the mutational enrichment of Iberian lineage was associated to genes encoding membrane proteins described with potential of interaction with external stimulus, as well as to genes with impact on the metabolism of the pathogen. However, future work should be done to better understand the metabolic consequences of these disarrangements and their impact in pathogenicity or virulence in a wide range of hosts, including man.

This genomic specialization and local adaptation of strains establish an Iberian ecovar, raising an important question regarding the mechanisms responsible for putative tropisms as response to adaptation to a specific host and/or pathobiological conditions.

5. Supplementary material

Supplementary Table S4.2.1. List of mutations disclosed in the comparative genomic analysis of 23 *Brucella* genomes and *Ochrobactrum anthropi* using as reference the annotated sequence of strain *B. suis* ATCC 23445. (A) List of SNPs; (B) List of INDELs; (C) List of INDELs differentiative of *B. suis* biovar 2 Iberian ecovar. Data can be seen at http://www.mediafire.com/file/2mm9zdwo7lyovm2/Supplementary_Table_S4.2.1_A-C.xlsx.

Supplementary Figure S4.2.1. Distribution of SNPs along the genome (SNPs per 0,2 Mb). Data can be seen at http://www.mediafire.com/file/3m9xsgpsfbf17tx/Supp.Figure_S4.2.1_SNPs_distribution.pdf.

Supplementary Table S4.2.2. *Brucella* strains used to search for the specific INDELs of Iberian Ecovar.

6. References

1. **Alton, G.C., Jones, L.M., Angus, R.D. & Verger, J.M. (1988)**. Techniques for the Brucellosis Laboratory. Institute National de la Recherche Agronomique, Paris, France.
2. **Ashkenazy, H., Cohen, O., Pupko, T. & Huchon, D. (2014)**. INDEL reliability in INDEL-Based Phylogenetic Inference. Genome Biology and Evolution, 6: 3199–3209.
3. **Audic, S., Lescot, M., Claverie, J.M. & Scholz, H.C. (2009)**. *Brucella microti*: the genome sequence of an emerging pathogen. BMC Genomics, 10: 352. doi: 101186/1471-2164-10-352 2009.
4. **Bäumler, A. & Fang, F.C. (2013)**. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med, 3: a010041. doi: 10.1101/cshperspect.a010041.
5. **Cardozo, P.G., Macedo, G.C., Azevedo, V. & Oliveira, S.C. (2006)**. *Brucella* spp. noncanonical LPS: structure, biosynthesis, and interaction with host immune system. Microbial Cell Factories, 5: 13 doi:10.1186/1475-2859-5-13.
6. **Chain, P.S., Comerci, D.J., Tolmasky, M.E., Larimer, F.W., Malfatti, S.A., Vergez, L.M., Aguero, F., Land, M.L., Ugalde, R.A. & Garcia, E. (2005)**. Whole-genome analyses of speciation events in pathogenic Brucellae. Infection and Immunity, 73(12): 8353-61.
7. **Chun, J. & Rainey, F.A. (2014)**. Integrating genomics into the taxonomy and systematics of the Bacteria and *Archaea*. International Journal of Systematic and Evolutionary Microbiology, 64: 316–324. DOI 10.1099/ijs.0.054171-0.
8. **Cloeckaert, A., Verger, J.M., Grayon, M. & Grépinet, O. (1995)**. Restriction site polymorphism of the genes encoding the major 25 and 36 kDa outer-membrane proteins of *Brucella*. Microbiology, 141: 2111–2121.
9. **Duvnjak, S., Račić, I., Špičić, S., Zdelar-Tuk, M., Reil, I. & Cvetnić, Ž. (2015).** Characterisation of *Brucella suis* isolates from Southeast Europe by multi-*locus* variable-number tandem repeat analysis. Veterinary Microbiology, 180(1-2): 146-50.
10. **EFSA (2009).** Scientific Opinion of the Panel on Animal Health and Welfare (AHAW) on a request from the Commission on porcine brucellosis (*Brucella suis*). The EFSA Journal, 1144: 1–112.
11. **Ferrão-Beck, L., Cardoso, R., Muñoz, P.M., de Miguel, M.J., Albert, D., Ferreira, A.C., Marín, C.M., Thiébaud, M., Jacques, I., Grayon, M., Zygmunt, M.S., Garin-Bastuji, B., Blasco, J.M. & Sá, M.I. (2006)**. Development of a multiplex PCR assay for polymorphism analysis of *Brucella suis* biovars causing brucellosis in swine. Veterinary Microbiology, 115(1-3): 269-77.

12. **Ferooz, J. & Letesson, J.J. (2010)**. Morphological analysis of the sheathed flagellum of *Brucella melitensis*. BMC Res Notes, 3:333. doi: 10.1186/1756-0500-3-333.Ferreira AC, Dias R, Corrêa de Sá MI and Tenreiro R (2016). Whole-genome mapping reveals a large chromosomal inversion on Iberian *Brucella suis* biovar 2 strains. Veterinary Microbiology, 192: 220–225. doi:10.1016/j.vetmic.2016.07.024.
13. **Ferreira, A.C., Tenreiro, R., Dias, R. & Corrêa de Sá, M.I. (2016).** Whole-genome mapping reveals a large chromosomal inversion on Iberian *Brucella suis* biovar 2 strains. Veterinary Microbiology, 192: 220–225.
14. **Ferreira, A.C., Tenreiro, R., Corrêa de Sá, M.I. & Dias, R. (2014a)**. Complete Genome Sequences of Three Iberian *Brucella suis* Biovar 2 Strains Isolated from Wild Boars. Genome Announcementsment, 2(4) pii: e00618-14 doi: 101128/genomeA00618-14 2014.
15. **Ferreira, A.C., Tenreiro, R., Corrêa de Sá, M.I. & Dias, R. (2014b)**. Complete Genome Sequences of Two Central-European *Brucella suis* bv. 2 Haplotype 2c Strains Isolated from Wild Boars. Genome Announcements, 2(4) pii: e00686-14 doi: 101128/genomeA00686-14 2014.
16. **Ficht, T. (2010)**. *Brucella* taxonomy and evolution. Future Microbiology, 5: 859–866.
17. **Forward, J.A., Behrendt, M.C., Wyborn, N.R., Cross, R. & Kelly, D.J. (1997)**. TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. Journal of Bacteriology, 179(17): 5482-93.
18. **Foster, J.T., Price, L.B., Beckstrom-Sternberg, S.M., Pearson, T., Brown, W.D., Kiesling, D.M., Allen, *C.A.*, Liu, C.M., Beckstrom-Sternberg, J., Roberto, F.F. & Keim, P. (2012)**. Genotyping of *Brucella* species using clade specific SNPs. BMC Microbiology, 12: 110.
19. **Foster, J.T., Beckstrom-Sternberg, S.M., Pearson, T., Beckstrom-Sternberg, J.S., Chain, P.S., Roberto, F.F., Hnath, J., Brettin, T. & Keim, P. (2009)**. Whole-genome-based phylogeny and divergence of the genus *Brucella*. Journal of Bacteriology, 191(8): 2864-70. doi: 101128/JB01581-08.
20. **Fretin, D., Fauconnier, A., Köhler, S., Halling, S., Léonard, S., Nijskens, C., Ferooz, J., Lestrate, P., Delrue, R.M., Danese, I., Vandenhaute, J., Tibor, A., DeBolle, X. & Letesson, J.J. (2005)**. The sheathed flagellum of *Brucella melitensis* is involved in persistence in a murine model of infection. Cell Microbiology, 7(5):687-98.
21. **Garcia-Yoldi, D., Le Flèche, P., De Miguel, M.J., Muñoz, P.M., Blasco, J.M., Cvetnic, Z., Marin, C.M., Vergnaud, G. & López-Goñi, I. (2007)**. Comparison of multiple-*locus* variable-number tandem-repeat analysis with other PCR-based methods for typing *Brucella suis* isolates. Journal of Clinical Microbiology, 45(12): 4070-2.
22. **Georgiades, K. & Raoult, D. (2010)**. Defining pathogenic bacterial species in the genomic era. Frontiers in Microbiology, 1:151.
23. **Godfroid, J., Garin-Bastuji, B., Saegerman, C. & Blasco, J.M. (2013)**. Brucellosis in terrestrial wildlife. Revue Scientifique et Technique-Office International des Epizooties, 32(1): 27-42.
24. **Halling, S.M., Peterson-Burch, B.D., Bricker, B.J., Zuerner, R.L., Qing, Z., Li, L.L., Kapur, V., Alt, D.P. & Olsen, S.C. (2005)**. Completion of the genome sequence of *Brucella abortus* and comparison to the highly similar genomes of *Brucella melitensis* and *Brucella suis*. Journal of Bacteriology, 187(8):2715-26 2005.
25. **Haiko, J. & Westerlund-Wikström, B. (2013)**. The role of the bacterial flagellum in adhesion and virulence. Biology (Basel), 2(4):1242-67. doi: 10.3390/biology2041242.
26. **Hughes, D. (2000)**. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. Genome Biology, 1(6): reviews0006.1–reviews0006.8.
27. **Jumas-Bilak, E., Michaux-Charachon, S., Bourg, G., O'Callaghan, D. & Ramuz, M. (1998)**. Differences in chromosome number and genome rearrangements in the genus *Brucella*. Molecular Microbiology, 27(1): 99–106.
28. **Kreizinger, Z., Foster, J.T., Ronaim Z., Sulyok, K.M., Wehmann, E., Janosi, S. & Gyuranecz, M. (2014)**. Genetic relatedness of *Brucella suis* biovar 2 isolates from hares, wild boars and domestic pigs. Veterinary Microbiology, 172(3-4): 492-8. doi: 101016/jvetmic201405031.
29. **Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C. & Salzberg, S.L. (2004)**. Versatile and open software for comparing large genomes. Genome Biology, 5(2): R12.

30. **Le Flèche, P., Jacques, I., Grayon, M., Al Dahouk, S., Bouchon, P., Denoeud, F., Nöckler, K., Neubauer, H., Guilloteau, L.A. & Vergnaud, G. (2006)**. Evaluation and selection of tandem repeat *loci* for a *Brucella* MLVA typing assay. BMC Microbiology, 6: 9.
31. **Li, Z.J., Cui, B.Y., Chen, H., Chen, J.D., Zhao, H.Y., Piao, D.R., Jiang, H., Zhang, L., Tang, X., Ke, C.W., Yao, Z. & Tian, G.Z. (2013)**. Molecular typing of *Brucella suis* collected from 1960s to 2010s in China by MLVA and PFGE. Biomedical and Environmental Sciences, 26(6): 504-8.
32. **Luan, P-T., Ryder, O.A., Davis, H., Zhang, Y-P. & Yu, L. (2013)**. Incorporating INDELs as phylogenetic characters: Impact for interfamilial relationships within Arctoidea (Mammalia: Carnivora). Molecular Phylogenetics and Evolution, 66: 748–756.
33. **Michaux-Charachon, S., Bourg, G., Jumas-Bilak, E., Guigue-Talet, P., Allardet-Servent, A., O'Callaghan. D. & Ramuz, M. (1997)**. Genome structure and phylogeny in the genus *Brucella*. Journal of Bacteriology, 179(10): 3244-9.
34. **Muñoz, P.M., Boadella, M., Arnal, M., de Miguel, M.J., Revilla, M., Martínez, D., Vicente, J., Acevedo, P., Oleaga, A., Ruiz-Fons, F., Marín, C.M., Prieto, J.M., de la Fuente, J., Barral, M., Barberán, M., de Luco, D.F., Blasco, J.M. & Gortázar, C. (2010)**. Spatial distribution and risk factors of Brucellosis in Iberian wild ungulates. BMC Infectious Diseases, 10: 46. doi: 101186/1471-2334-10-46 2010.
35. **Olsen, S.C. & Palmer, M.V. (2014).** Advancement of knowledge of *Brucella* over the past 50 years. Veterinary Pathology, 51(6): 1076-89. doi: 10.1177/0300985814540545.
36. **Paulsen, I.T., Seshadri, R., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Read, T.D., Dodson, R.J., Umayam, L., Brinkac, L.M., Beanan, M.J., Daugherty, S.C., Deboy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Nelson, W.C., Ayodeji, B., Kraul, M., Shetty, J., Malek, J., Van Aken, S.E., Riedmuller, S., Tettelin, H., Gill, S.R., White, O., Salzberg, S.L., Hoover, D.L., Lindler, L.E., Halling, S.M., Boyle, S.M. & Fraser, C.M. (2002)**. The *Brucella suis* genome reveals fundamental similarities between animal and plant pathogens and symbionts. Proceedings of the National Academy of Sciences of the United States of America, 99(20): 13148-53.
37. **Posadas, D.M., Ruiz-Ranwez, V., Bonomi, H.R., Martín, F.A. & Zorreguieta, A. (2012)**. BmaC, a novel autotransporter of *Brucella suis*, is involved in bacterial adhesion to host cells. Cell Microbiology, 14(6): 965-82. doi: 10.1111/j.1462-5822.2012.01771.x.
38. **Redelings, B.D. & Suchard, M.A. (2007)**. Incorporating INDEL information into phylogeny estimation for rapidly emerging pathogens. BMC Evolutionary Biology, 7:40.
39. **Ruiz-Ranwez, V., Posadas, D.M., Estein, S.M., Abdian, P.L., Martin, F.A. & Zorreguieta, A. (2013a)**. The BtaF trimeric autotransporter of *Brucella suis* is involved in attachment to various surfaces, resistance to serum and virulence. PLoS One, 8(11): e79770. doi: 10.1371/journal.pone.0079770.
40. **Ruiz-Ranwez, V., Posadas, D.M., Van der Henst, C., Estein, S.M., Arocena, G.M., Abdian, P.L., Martín, F.A., Sieira, R., De Bolle, X. & Zorreguieta, A. (2013b)**. BtaE, an adhesin that belongs to the trimeric autotransporter family, is required for full virulence and defines a specific adhesive pole of *Brucella suis*. Infection and Immunity, 81(3):996-1007. doi: 10.1128/IAI.01241-12.
41. **Sankarasubramanian, J., Vishnu, U.S., Gunasekaran, P. & Rajendhran, J. (2016)**. A genome-wide SNP-based phylogenetic analysis distinguishes different biovars of *Brucella suis*. Infection, Genetics and Evolution, 41: 213-7. doi: 10.1016/j.meegid.2016.04.012.
42. **Saitou, N. & Nei, M. (1987)**. The Neighbor-Joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4: 406-425.
43. **Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. (2013)**. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution, 30: 2725-2729.
44. **Tsolis, R.M., Seshadri, R., Santos, R.L., Sangari, F.J., Lobo, J.M., de Jong, M.F., Ren, Q., Myers, G., Brinkac, L.M., Nelson, W.C., Deboy, R.T., Angiuoli, S., Khouri, H., Dimitrov, G., Robinson, J.R., Mulligan, S., Walker, R.L., Elzer, P.E., Hassan, K.A. & Paulsen, I.T. (2009)**. Genome degradation in *Brucella ovis* corresponds with narrowing of its host range and tissue tropism. PLoS One, 4(5): e5519. doi: 10.1371/journal.pone.0005519.

45. **Vizcaíno, N., Verger, J.M., Grayon, M., Zygmunt, M.S. & Cloeckaert, A. (1997)**. DNA polymorphism at the omp-31 *locus* of *Brucella* spp.: evidence for a large deletion in *Brucella abortus*, and other species-specific markers. Microbiology, 143: 2913–2921.
46. **Wattam, A.R., Williams, K.P., Snyder, E.E., Almeida, N.F. Jr, Shukla, M., Dickerman, A.W., Crasta, O.R., Kenyon, R., Lu, J., Shallom, J.M., Yoo, H., Ficht, T.A., Tsolis, R.M., Munk, C., Tapia, R., Han, C.S., Detter, J.C., Bruce, D., Brettin, T.S., Sobral, B.W., Boyle, S.M. & Setubal, J.C. (2009)**. Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle. Journal of Bacteriology, 191(11): 3569-79. doi: 10.1128/JB.01767-08.
47. **Whatmore, A.M., Perrett, L.L. & MacMillan, A.P. (2007)**. Characterization of the genetic diversity of *Brucella* by multi*locus* sequencing. BMC Microbiology, 7: 34.
48. **Zygmunt, M.S., Vizcaíno, N., Cloeckaert, A., Grilló, M.J., Moriyón, I. & López-Goñi, I. (2014)**. Mutants in the lipopolysaccharide of *Brucella ovis* are attenuated and protect against *B. ovis* infection in mice. Veterinary Research, 45: 72. doi: 10.1186/s13567-014-0072-0.

Supplementary Table S4.2.2. *Brucella* strains used to search for the specific INDELs of Iberian ecovar.

Strain	Biovar	Host	Country	SI79506	SI1356057	SI1423448	SI1627485	SI2041144	SI2603410	LI3234619
REF 1330	1	Swine	USA	598	714	297	180	228	508	2,170
REF ATCC 23445	2	Hare	Denmark	938	853	297	180	228	1,352	2,170
REF 686	3	Swine	USA	598	714	297	180	228	508	2,170
REF 40	4	Reindeer	Former USSR	598	714	297	180	228	508	2,170
REF 513	5	Wild rodent	Former USSR	598	765	297	180	228	508	2,170
REF 16M	1	Goat	USA	598	765	297	180	228	508	2,170
REF 63/9	2	Goat	Turkey	598	765	297	180	228	508	2,170
REF Ether	3	Goat	Italia	598	765	297	180	228	508	2,170
REF 544	1	Cattle	England	2,293	765	297	180	228	508	2,170
REF 2308	1	Cattle	USA	2,293	765	297	180	228	508	2,170
S19	1	Cattle	USA	598	765	297	180	228	508	2,170
REF 86/8/59	2	Cattle	England	598	765	297	180	228	508	2,170
REF Tulya	3	Human	Uganda	598	765	297	180	228	508	2,170
REF 292	4	Cattle	USA	MB	765	297	180	228	508	2,170
REF B3196	5	Cattle	USA	MB	765	297	180	228	508	2,170
REF 870	6	Cattle	USA	MB	765	297	180	228	508	2,170
REF C68	9	Cattle	USA	MB	765	297	180	228	508	2,170
REF 63/290	na	Sheep	Australia	598	765	297	180	228	508	2,170
REF 5K33	na	Dog	USA	598	714	297	180	228	508	2,170
REF B1/94	na	Dolphin	Scotland	598	853	297	180	228	508	2,170
B2/94	na	Seal	Scotland	598	765	297	180	228	508	2,170
CCM 4915	na	Wild rodent	Rep.Czech	1,276	765	297	180	228	508	2,170
Field Isolates										
INRA	1	Swine	France	598	714	297	180	228	508	2,170
96/ 1646-01	1	Swine	France	598	714	297	180	228	508	2,170
01/ 5744	1	Swine	Polynesia	598	714	297	180	228	508	2,170
03/ 2067-203	1	Swine	France	598	714	297	180	228	508	2,170
03/ 3081-2	1	Goat	Croatia	598	714	297	180	228	508	2,170
04/ 115	1	Hare	France	598	714	297	180	228	508	2,170
04/ 2987	1	Human	France	598	714	297	180	228	508	2,170
04/ 3025-3	1	Swine	Croatia	598	714	297	180	228	508	2,170
64/24	1	Swine	USA	598	714	297	180	228	508	2,170
F1/04	1	Human	Netherlands	598	714	297	180	228	508	2,170
92/29	1	Human	Mexico	598	714	297	180	228	508	2,170
98/ 7296-4204	2	Hare	France	938	853	297	180	228	1,352	2,170
A183	2	Wild boar	Germany	938	853	297	180	228	508	2,170
A196	2	Wild boar	Germany	938	853	297	180	228	508	2,170
04RB0377	2	Wild boar	Germany	938	853	297	180	228	508	2,170
05RB0007	2	Wild boar	Germany	938	853	297	180	228	508	2,170
05RB1442	2	Hare	Germany	938	853	297	180	228	508	2,170
74/11	2	Unknown	Denmark	938	853	297	180	228	1,352	2,170
92/ 11580-4528	2	Hare	France	938	853	297	180	228	1,352	2,170
92/ 13000	2	Hare	France	938	853	297	180	228	1,352	2,170
98/ 6335	2	Swine	France	938	853	297	180	228	1,352	2,170
00/ 9182	2	Hare	France	938	853	297	180	228	1,352	2,170
96/ 9635	2	Swine	France	938	853	297	180	228	1,352	2,170
97/ 4924-10	2	Swine	France	938	853	297	180	228	1,352	2,170
97/ 9757	2	Swine	France	938	853	297	180	228	1,352	2,170
00/ 4898	2	Bovine	France	938	853	297	180	228	1,352	2,170

Strain	Biovar	Host	Country	SI79506	SI1356057	SI1423448	SI1627485	SI2041144	SI2603410	LI3234619
03/ 1483-8	2	Wild boar	France	938	853	297	180	228	1,352	2,170
04/ 770	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
04/ 1918-1	2	Wild boar	Switzerland	938	853	297	180	228	1,352	2,170
04/ 3025-1	2	Swine	Croatia	938	853	297	180	228	1,352	2,170
S120	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
COSA 13	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
MASA 07	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
PY69	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
RATE S5-11	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C13B6	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C2B11	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C9B4	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C6B1	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C8B3	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C9B3	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C13B1	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C13B4	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C4B6	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C5B5	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C3B3	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
C11B4	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
S275	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
0111602/4+9	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
0111602/3+8	2	Wild boar	Belgium	938	853	297	180	228	1,352	2,170
It1	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
It2	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
It3	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
It4	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
It5	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
It6	2	Wild boar	Italy	938	853	297	180	228	1,352	2,170
Bs364CITA	**2**	**Wild boar**	**Spain**	**938**	**853**	**297**	**180**	**228**	**1,352**	**2,170**
S-365	2	Wild boar	Spain	938	853	297	180	228	1,352	2,170
Bs396CITA	**2**	**Wild boar**	**Spain**	**938**	**853**	**297**	**180**	**228**	**1,352**	**2,170**
7676/00	2	Swine	Portugal	260	765	297	129	228	508	174
438/00	2	Swine	Portugal	260	765	297	129	228	508	174
5346(1)/00	2	Swine	Portugal	260	765	297	129	228	508	174
5346 (16)/00	2	Swine	Portugal	260	765	297	129	228	508	174
5346(3)/00	2	Swine	Portugal	260	765	297	129	228	508	174
1967/00	2	Swine	Portugal	260	765	297	129	228	508	174
6552(3)/00	2	Swine	Portugal	260	765	297	129	228	508	174
6552(8)/00	2	Swine	Portugal	260	765	297	129	228	508	174
6552(2)/00	2	Swine	Portugal	260	765	297	129	228	508	174
6552(10)/00	2	Swine	Portugal	260	765	297	129	228	508	174
S1	2	Swine	Spain	260	765	297	129	228	508	174
S2	2	Swine	Spain	260	765	297	129	228	508	174
S3	2	Swine	Spain	260	765	297	129	228	508	174
S4	2	Swine	Spain	260	765	297	129	228	508	174
S22	2	Swine	Spain	260	765	297	129	228	508	174
S32	2	Swine	Spain	260	765	297	129	228	508	174
S34	2	Swine	Spain	260	765	297	129	228	508	174
S6	2	Swine	Spain	260	765	297	129	228	508	174
S21	2	Swine	Spain	260	765	297	129	228	508	174
2739 (3)/02	2	Swine	Portugal	260	765	297	129	228	508	174

Strain	Biovar	Host	Country	SI79506	SI1356057	SI1423448	SI1627485	SI2041144	SI2603410	LI3234619
2739(8)/02	2	Swine	Portugal	260	765	297	129	228	508	174
2739(9)/02	2	Swine	Portugal	260	765	297	129	228	508	174
5414(12)/03	2	Swine	Portugal	260	765	297	129	228	508	174
5414(13)/03	2	Swine	Portugal	260	765	297	129	228	508	174
5414(2)/03	2	Swine	Portugal	260	765	297	129	228	508	174
5414(13)/03	2	Swine	Portugal	260	765	297	129	228	508	174
468/00	2	Swine	Portugal	260	765	297	129	228	508	174
1262/00	2	Swine	Portugal	260	765	297	129	228	508	174
S/29/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/35/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/39/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/41/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/43/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/50/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/57/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/64/2000	2	Swine	Portugal	260	765	297	129	228	508	174
S/73/2000	2	Swine	Portugal	260	765	297	129	228	508	174
00/ 5952-1	2	Swine	Portugal	260	765	297	129	228	508	174
Bs146	2	Swine	Spain	260	765	297	129	228	508	174
Bs147	2	Swine	Spain	260	765	297	129	228	508	174
UK1	2	Swine	Unknown	260	765	297	129	228	508	174
2454(J10)/ 08	2	Wild boar	Portugal	260	765	297	129	228	508	174
4498 (J1)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
4498(J9)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
4498(J4)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
4498(J6)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
9789 (J9)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
9789 (J10)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
9789 (Feto)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
19122 (J1)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
17888 (J3)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
21346 (J2)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
22497 (475)/08	2	Sheep	Portugal	260	765	297	129	228	508	174
22498 (697)/08	2	Sheep	Portugal	260	765	297	129	228	508	174
46685 (15)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
46685 (20)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
44401(3)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
44406(3)/08	2	Wild boar	Portugal	260	765	297	129	228	508	174
3115 (9)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
3115 (13)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
1344 (2)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
1344 (3)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
3478 (3)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
14256 (40)/09	2	Swine	Portugal	260	765	297	129	228	508	174
14256 (72)/09	2	Swine	Portugal	260	765	297	129	228	508	174
231 (1)/09	2	Swine	Portugal	260	765	297	129	228	508	174
21566 (94)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
44821(81-f)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
44821 (86-b)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
44821 (79-f)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
1989 (112-Fa)/10	2	Wild boar	Portugal	260	765	297	129	228	508	174
PT09172	**2**	**Wild boar**	**Portugal**	**260**	**765**	**297**	**129**	**228**	**508**	**174**
44821 (121)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174

Strain	Biovar	Host	Country	SI79506	SI1356057	SI1423448	SI1627485	SI2041144	SI2603410	LI3234619
44821 (122)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
44821 (123)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
S-384	2	Swine	Spain	260	765	297	129	228	508	174
4187(7)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
4478(4)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
4647(1)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
4193(1)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
41932)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
4193(9)/09	2	Wild boar	Portugal	260	765	297	129	228	508	174
3514(2)/10	2	Wild boar	Portugal	260	765	297	129	228	508	174
3515(6)/10	2	Wild boar	Portugal	260	765	297	129	228	508	174
4215 (20)/01	2	Swine	Portugal	260	765	297	129	228	508	174
9H	2	Swine	Portugal	260	765	297	129	228	508	174
S27	2	Swine	Spain	260	765	218	129	171	508	174
S12	2	Swine	Spain	260	765	218	129	171	508	174
S13	2	Swine	Spain	260	765	218	129	171	508	174
Bs144	2	Swine	Spain	260	765	218	129	171	508	174
Bs145	2	Swine	Spain	260	765	218	129	171	508	174
J2A-F/CM/08	2	Wild boar	Portugal	260	765	218	129	171	508	174
45014 (4)/08	2	Wild boar	Portugal	260	765	218	129	171	508	174
2948 (1)/09	2	Wild boar	Portugal	260	765	218	129	171	508	174
2948 (5)/09	2	Wild boar	Portugal	260	765	218	129	171	508	174
PT09143	**2**	**Wild boar**	**Portugal**	**260**	**765**	**218**	**129**	**171**	**508**	**174**
2948 (34)/09	2	Wild boar	Portugal	260	765	218	129	171	508	174
4189 (1)/09	2	Wild boar	Portugal	260	765	218	129	171	508	174
4189 (2)/09	2	Wild boar	Portugal	260	765	218	129	171	508	174
Bs143CITA	**2**	**Wild boar**	**Spain**	**260**	**765**	**218**	**129**	**171**	**508**	**174**
S-145 (PN-II)	2	Wild boar	Spain	260	765	218	129	171	508	174
4477 (4)/09	2	Wild boar	Portugal	260	765	218	129	171	508	174
8605(2)/10	2	Wild boar	Portugal	260	765	218	129	171	508	174
8605(9)/10	2	Wild boar	Portugal	260	765	218	129	171	508	174
UK 6/04	3	Human	India	598	714	297	180	228	508	2,170
63/252	4	Caribou	Alaska	598	714	297	180	228	508	2,170
63/198	4	Reindeer	Poland	598	714	297	180	228	508	2,170
63/202	4	Reindeer	Poland	598	714	297	180	228	508	2,170
63/219	4	Reindeer	Poland	598	714	297	180	228	508	2,170

Chapter 5

Concluding remarks and future perspectives

Concluding remarks

The brucellosis infection in pigs, in Portugal as well as in other European countries, has been practically silent for years. The control standards for this animal host only concerned the screening of breeding animals for artificial insemination and for export. However, in recent years there has been an increase in the number of cases of *B. suis* biovar 2 in pigs with abortions and positive serology reactions, especially in extensive farms. This has major economic impact since it hampers the free movement of animals and the commerce of their products. Currently, there is a great lack of knowledge about the epidemiological aspects of this infection, the population structure of *B. suis* biovar 2 isolates, its geographic distribution and host specificity. Direct interaction between infected wildlife and livestock at a local interface may be the most important driver for periodic spillback transmission of *Brucella*. Therefore, there is a need for critical deliberation of its epidemiology, pathogenesis and diagnosis to improve prevention and management at local, regional, national and global levels.

As reviewed in **Chapter 1**, most studies on *Brucella* spp. are based on *B. abortus, B. melitensis* and *B. suis* (biovar 1), which are the most important species regarding its worldwide distribution and virulence for Man. Nevertheless, swine brucellosis due to *B. suis* biovar 2 is an emerging disease in Europe associated with extensive swine farms and high density of infected wild boars. Although the few work developed to understand the epidemiological relationships and the phylogeny of this pathogen, a considerable intraspecies genetic diversity has already been observed. One of the main questions addressed in this thesis is related with the characterization of differences between *B. suis* biovar 2 Iberian strains and those from Central Europe, and further unveil strain specific genomic variations associated with geographic origins. Moreover, it is unknown which molecular or physiological mechanisms are responsible for the loss of virulence of this biovar to humans. Further sequencing of the genomes of Iberian field strains seemed thus crucial to disclose the structure and distinctive features of *B. suis* pan-genome, as well as its evolutive history, host preference and geographic partitioning.

The “gold standard” for the diagnosis of brucellosis is the bacterial isolation. However, the slow growth of *Brucella* species, combined with their requirement for highly nutritious media means that selective reagents must be incorporated to prevent overgrowth of contaminant organisms that can be inhibitory to some *Brucella* strains. Additionally, the use of increasingly powerful genotyping tools for the characterization of pathogens has become a standard component of infectious disease surveillance and outbreak investigations.

In Portugal, the Laboratório Nacional de Referência de Saúde Animal from INIAV, is the Nacional Reference Laboratory for Animal Brucellosis. Nevertheless, the diagnosis is mainly focused on bovine and caprine brucellosis and isolation of *B. abortus* and *B. melitensis*. Although *B. suis* infection has been previously demonstrated in swine and wild boars, few *B. suis* isolates were obtained and their molecular characterization was limited to few studies. In **Chapter 2**, a selective medium, LNIV-M, was developed and evaluated for the primary isolation of *B. suis* in comparison with other media. Accordingly, LNIV-M was formulated using less inhibitory antibiotics, improving the quality of bacteriological diagnostic and enhancing the number of *B. suis* isolates from wild boars and swine from different regions of Portugal. The studies presented in **subchapters 2.1 and 2.2** aimed to extend our understanding on the prevalence and molecular epidemiology of *B. suis* biovar 2 in Portugal. Since knowledge of predominant circulating strains is a prerequisite for any epidemiological study, accurate molecular typing procedures were applied to a collection of *B. suis* isolates, that included isolates from Portugal as well as isolates from other European countries. By using Suis-ladder multiplex PCR and PCR-RFLP analysis of *omp2a*, *omp2b* and *omp31* genes, five haplotypes were identified among biovar 2 isolates, with specific haplotypes restricted to Portugal and Spain (2d and 2e) and haplotypes 2a, 2b and 2c widespread in Europe (except Portugal). The genetic diversity of *B. suis* population was evaluated using MLVA based on 16 genetic markers (MLVA-16). The analysis grouped biovar 2 isolates in two clusters according to their geographic origins and haplotypes, defining the Iberian (Portugal and Spain) and the Central-European clonal lineages. Furthermore, an extended analysis was performed using a subset of 11 markers and publicly available data for 350 additional strains, revealing a high genetic divergence amongst *B. suis* strains based on their hosts and highlighting the close relationship between strains from swine, wild boars and hares. Beyond corroborating the existence of Iberian and Central-European biovar 2 clonal lineages and pointing to the evolution of biovar 2 Iberian clonal lineage from Central-European by an allopatric speciation event, an ongoing colonization of Iberian Peninsula with specific MLVA-11 genotypes was also observed.

In **Chapter 3**, optical maps were used to compare closely related *B. suis* biovar 2 strains. Optical mapping is a technology able to quickly generate high resolution ordered whole-genome restriction maps of bacteria, being a proven approach to search for diversity among bacterial isolates. We performed the *BamH* I whole-genome optical maps of five *B. suis* biovar 2 field strains, isolated from wild boars in Portugal and Spain, representative of the two clonal lineages circulating in Iberian Peninsula as well as of the reference strain *B. suis* biovar 2 ATCC 23445 (Central-European lineage, Denmark). Each strain showed a distinct, highly individual configuration of 228–231 *BamH* I fragments, distributed in the two chromosomes. Nevertheless, a low divergence was globally observed in chromosome II (1.6%) relatively to chromosome I (2.4%). Optical mapping also disclosed genomic

events associated with *B. suis* strains in chromosome I, namely one INDEL (3.5 kb) and one large inversion (944 kb). The INDEL was found to be specific of the reference strain ATCC 23445 and the large inversion was shown to be an exclusive genomic marker of the Iberian clonal lineage of biovar 2.

The upmost goal of the work described here was to obtain the complete and annotated genome sequences of five *B. suis* biovar 2 Iberian field strains in order to better understand the mechanisms of evolution and specialization of Iberian lineages (**Chapter 4**). In **Subchapter 4.1**, the full genomes of five *B. suis* biovar 2 field strains were promptly obtained using a combination of Illumina, Sanger and optical mapping technologies. Besides validating the existence of an inversion in biovar 2 Iberian clonal lineage, and thus reinforcing the plasticity of *Brucella* genomes, the optical maps defined unique genome landmarks in each of the strains and demonstrated the ability of this method to determine the relative placement and orientation of sequence fragments produced during the assembly process, being very useful for closing genomes. The comparative genomic analysis of *B. suis* biovar 2 together with publicly available *Brucella* spp. genomes using three whole-genome approaches (**Subchapter 4.2**) showed that brucellae are genetically highly conserved organisms that have apparently slow evolutionary rates. Nevertheless, it was observed that *B. suis* biovar 2 strains from Iberian clonal lineage are differentiated from those from Central-European clonal lineage not only by the presence of one large inversion in Chr I but also by a number of specific SNPs and biovar-, haplotype- and strain-specific insertion-deletion (INDELs) events that could explain differences in virulence and host specificities. It was observed that most of the discriminative mutations were associated to membrane related molecules (29%) and enzymes involved in catabolism process (20%). Moreover, our results suggest the genomic specialization of *B. suis* biovar 2 on the Iberian Peninsula to be independent of a specific genomic event(s), raising an important question regarding the mechanisms responsible for putative tropisms as response to adaptation to a specific host and/or pathobiological conditions. In nature, brucellae are isolated as they multiply within their preferential host. This has led to the suggestion that the different strains of *Brucella* are clones that have co-evolved independently with their mammalian hosts. One of the permanent mysteries of *Brucella* remains the basis of their host specificity in the face of the relatively minor genetic variation between species. Identification of genetic differences provides a realistic framework for examining the biological basis of host specificity. In addition, seven polymorphic regions with epidemiological marker potential (*e.g.* exclusive of *B. suis* Iberian clonal lineage) were identified and evaluated using the target-PCR procedures established in this work.

In conclusion, it was demonstrated that two *B. suis* biovar 2 clonal lineages circulates in Europe and a genomic specialization and local adaptation of strains in Iberian Peninsula established an Iberian ecovar. However, it is suggested that the mechanisms for this genomic specialization were

independent of the chromosomal rearrangement events observed. Furthermore, the results obtained in this work point to the evolution of this biovar 2 Iberian ecovar from the Central-European clonal lineage by an allopatric speciation event.

Future perspectives

At the end of the work here described and discussed, particular questions remain to be answered. One refers to the biological significance of the chromosomal inversion detected in Iberian strains. The inversion disrupted a TRAP dicarboxylate transporter, DctM subunit, and an integral membrane protein TerC. The *dctM* gene is well conserved in *Brucella* spp. except in strains from Iberian clonal lineage. Additional investigation is required to understand and analyze the potential impact of this inversion in the metabolism of *B. suis* biovar 2 strains belonging to the Iberian ecovar.

In relation with the polymorphisms found on genes coding for flagellar and outer membrane proteins described as belonging to putative autotransporter family, it remains to be clarified and proved if those changes would influence the formation of a functional flagellum and consequently affect several functions such as protein export or adhesion. Likewise, future work should be done to realize if the inactivation of specific autotransporter proteins would contribute to host-pathogen interactions thus influencing *B. suis* biovar 2 virulence. In fact, no INDELs were found to affect genes known to participate in virulence and it is known that the genomes of *Brucella* species have low polymorphisms. This can be an advantage because if polymorphisms are relatively rare, the ones we identified are likely to be important to understand *B. suis* biovar 2 molecular tropism for wild boars or pigs.

Another relevant topic concerns the results that suggest the ancestral origin of *Brucella* genus to be related to *B. suis* biovar 2, supporting its use in the design of novel vaccines based on targets that are universally conserved across all major pathogenic *Brucella* species. With the advent of whole-genome sequencing and advances in bioinformatics, the vaccinology field is radically changed. Control and eradication of brucellosis require sanitary programs, based on the screening of animals and diagnosis of the disease, and, as a prophylactic measure, the use of vaccines. The availability of an efficacious vaccine is thus an important future objective that should be considered as a priority.